Sharp rate for the dual quantization problem

Gilles Pagès, Benedikt Wilbertz

Sharp rate for the dual quantization problem

Gilles Pagès* Benedikt Wilbertz†

January 24, 2015

Abstract

In this paper we establish the sharp rate of the optimal dual quantization problem. The notion of dual quantization was recently introduced in [12], where it has been shown that, at least in a Euclidean setting, dual quantizers are based on a Delaunay triangulation, the dual counterpart of the Voronoi tessellation on which "regular" quantization relies. Moreover, this new approach shares an intrinsic stationarity property, which makes it very valuable for numerical applications.

We establish in this paper the counterpart for dual quantization of the celebrated Zador theorem, which describes the sharp asymptotics for the quantization error when the quantizer size tends to infinity. On the way we establish an extension of the so-called Pierce Lemma by a random quantization argument. Numerical results confirm our choices.

Keywords: quantization, quantization rate, Zador's Theorem, Pierce's Lemma, dual quantization, Delaunay triangulation, random quantization.

1 Introduction

Starting with [11] and continued in [12], we introduced a new notion of vector quantization called *dual quantization* (or *Delaunay quantization* in a Euclidean framework). We developed in [10] some first applications towards the design of numerical schemes for multi-dimensional optimal stopping and stochastic control problems arising in Finance (see also [1]). In general, the principle of dual quantization consists of mapping an $\mathbb{R}^d$-valued random vector (r.v.) onto a non-empty finite subset (or *grid*) $\Gamma \subset \mathbb{R}^d$ using an appropriate random splitting operator $\mathcal{J}_\Gamma : \Omega_0 \times \mathbb{R}^d \to \Gamma$ (defined on an exogenous probability space $(\Omega_0, \mathcal{S}_0, \mathbb{P}_0)$) which satisfies the intrinsic *stationarity property*

$$\forall\, \xi \in \operatorname{conv}(\Gamma), \qquad \mathbb{E}_{\mathbb{P}_0}(\mathcal{J}_\Gamma(\xi)) = \int_{\Omega_0} \mathcal{J}_\Gamma(\omega_0, \xi)\, \mathbb{P}_0(d\omega_0) = \xi, \tag{1}$$

where $\operatorname{conv}(\Gamma)$ denotes the convex hull of Γ in $\mathbb{R}^d$. Every r.v. $X : (\Omega, \mathcal{S}, \mathbb{P}) \to \operatorname{conv}(\Gamma)$ defined on a probability space can be canonically extended to $(\Omega_0 \times \Omega, \mathcal{S}_0 \otimes \mathcal{S}, \mathbb{P}_0 \otimes \mathbb{P})$ in order to define *dual quantization* induced by Γ as

$$\widehat{X}^{\Gamma,dual}(\omega_0, \omega) = \mathcal{J}_\Gamma(\omega_0, X(\omega)).$$

As a specific feature inherited from (1), it always satisfies the *dual* or *reverse stationary property*

$$\mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_0}(\mathcal{J}_\Gamma(X) \,|\, X) = X.$$

*Laboratoire de Probabilités et Modèles aléatoires, UMR 7599, Université Paris 6, case 188, 4, pl. Jussieu, F-75252 Paris Cedex 05. E-mail: `gilles.pages@upmc.fr`

†Laboratoire de Probabilités et Modèles aléatoires, UMR 7599, Université Paris 6, case 188, 4, pl. Jussieu, F-75252 Paris Cedex 05. E-mail: `benedikt.wilbertz@gmx.de`

This can be compared to the more classical Voronoi framework where the Γ-quantization of X is defined from a Borel nearest neighbour projection Proj_Γ by

$$\widehat{X}^{\Gamma,vor}(\omega) = \mathrm{Proj}_\Gamma(X(\omega)).$$

The stationary property then reads: $\mathbb{E}(X \,|\, \widehat{X}^{\Gamma,vor}) = \widehat{X}^{\Gamma,vor}$, except that it holds only for grids which are critical points (typically local minima) of the so-called distortion function (see *e.g.* [5]) in a Euclidean framework.

To each quantization is corresponds a functional approximation operator: Voronoi quantization is related to the *stepwise constant functional approximation operator* $f \circ \mathrm{Proj}_\Gamma$ whereas dual quantization leads to an operator defined for every $\xi \in \mathrm{conv}(\Gamma)$ by

$$\mathbb{J}_\Gamma(f)(\xi) = \mathbb{E}_{\mathbb{P}_0}\big(f(J_\Gamma(\omega_0,\xi))\big) = \sum_{x\in\Gamma} f(x)\lambda_x(\xi), \tag{2}$$

where $\lambda_x(\xi) = \mathbb{P}_0(J_\Gamma(.,\xi) = x)$, $x \in \Gamma$, are barycentric "pseudo-coordinates" of ξ in Γ satisfying $\lambda_x(\xi) \in [0,1]$, $\sum_{x\in\Gamma}\lambda_x(\xi) = 1$ and $\sum_{x\in\Gamma}\lambda_x(\xi)x = \xi$. The operator $\mathbb{J}_\Gamma$ is an *interpolation* operator which turns out, under appropriate conditions, to be more regular (continuous and stepwise affine, see [10]) than the "Voronoi" one. It is shown in [12, 11, 10] how we can take advantage of this intrinsic stationary property to produce more accurate error bounds for the resulting *cubature formula*

$$\mathbb{E}_{\mathbb{P}}(f(\widetilde{X}^{\Gamma,dual})) = \mathbb{E}_{\mathbb{P}}(\mathbb{J}_\Gamma(f)(X)) = \mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_0}\big(f(J_\Gamma(\omega_0,\xi))\big) = \sum_{x\in\Gamma} w_x^{dual} f(x) \tag{3}$$

where $w_x^{dual} = \mathbb{E}_{\mathbb{P}}(\lambda_x(X)) = \mathbb{P}\otimes\mathbb{P}_0(J_\Gamma(\omega_0,X) = x)$, $x \in \Gamma$, regardless of any optimality property Γ with respect to $\mathbb{P}_X$. Typically, if $f \in \mathrm{Lip}(\mathbb{R}^d,\mathbb{R})$ (Lipschitz continuous function) with coefficient $[f]_{\mathrm{Lip}}$,

$$\begin{aligned}\Big|\mathbb{E}_{\mathbb{P}}f(X) - \mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_0}f(\widetilde{X}^{\Gamma,dual})\Big| &\le [f]_{\mathrm{Lip}}\big\|X - \widehat{X}^{\Gamma,dual}\big\|_{L^1(\mathbb{P}\otimes\mathbb{P}_0)}\\ &= [f]_{\mathrm{Lip}}\mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_0}\big(\|X - J_\Gamma(\omega_0,X)\|\big)\\ &= [f]_{\mathrm{Lip}}\mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_0}\big(\mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_0}(\|X - J_\Gamma(\omega_0,X)\| \,|\, X)\big)\end{aligned}$$

whereas, if f has Lipschitz continuous differential (the norm on $\mathbb{R}^d$ is denoted $\|\,.\,\|$), a second order Taylor expansion yields

$$\begin{aligned}\Big|\mathbb{E}_{\mathbb{P}}f(X) - \mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_0}f(\widetilde{X}^{\Gamma,dual})\Big| &\le \Big\|f(X) - \mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_0}\big(f(J_\Gamma(\omega_0,X)) \,|\, X\big)\Big\|_{L^1(\mathbb{P}\otimes\mathbb{P}_0)}\\ &\le [Df]_{\mathrm{Lip}}\mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_0}\big(\|X - J_\Gamma(\omega_0,X)\|^2\big)\\ &\le [Df]_{\mathrm{Lip}}\mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_0}\big(\mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_0}(\|X - J_\Gamma(\omega_0,X)\|^2 \,|\, X)\big)\end{aligned} \tag{4}$$

where $\mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_0}(\|X - J_\Gamma(\omega_0,X)\|^p \,|\, X) = \displaystyle\sum_{x\in\Gamma}\lambda_x(X)\|X - x\|^p = \mathbb{J}_\Gamma(\|.\|^2)(X),\ p = 1, 2.$

More generally, if one aims at approximating $\mathbb{E}\big(f(X) \,|\, g(Y)\big)$ by its dually quantized counterpart $\mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_0\otimes\mathbb{P}_1}\big(f(J_{\Gamma_X}(\omega_0,X)) \,|\, J_{\Gamma_Y}(\omega_1,Y)\big)$ (with obvious notations), it is also possible under natural additional assumptions to get error bounds based on both related dual quantization error moduli, see e.g. the proof (Step 2) of Proposition 2.1 in [10].

This suggests to investigate the properties and the asymptotic behaviour of the (Γ, L^p)-mean dual quantization error, $p \in (0,\infty)$, defined by

$$\Big\|X - \widehat{X}^{\Gamma,dual}\Big\|^p_{L^p(\mathbb{P}\otimes\mathbb{P}_0)} = \Big\|X - J_\Gamma(\omega_0,X)\Big\|^p_{L^p(\mathbb{P}\otimes\mathbb{P}_0)} = \mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_0}\Big(\mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_0}\big(\|X - J_\Gamma(\omega_0,X)\|^p \,|\, X\big)\Big)$$

so as to make it as small as possible. This program can be summed up in four phases:

– The first step is to minimize the above conditional expectation, *i.e.* $\mathbb{E}(\|\xi - J_\Gamma(\omega_0,\xi)\|^p)$ for every $\xi \in \mathrm{conv}(\Gamma)$, for a fixed grid Γ *i.e.* to determine the *best* splitting random operator J_Γ. In a regular quantization, this phase corresponds to showing that the nearest neighbour projection on Γ is the best projection on Γ.

– The second step is "optional" . It aims at finding grids which minimize the mean dual quantization error $\left\|X - J_\Gamma(\omega_0, X)\right\|_{L^p(\mathbb{P}\otimes\mathbb{P}_0)}$ *among all grids* Γ *whose convex hull contains the support of the distribution of* X or equivalently such that $\mathbb{P}(X \in \mathrm{conv}(\Gamma)) = 1$.

– The third step is to extend dual quantization to r.v.s X with unbounded support while the performances of the resulting cubature formula (see (4)), having in mind that the stationarity can no longer holds.

The first two steps have been already solved in [12]. We discuss in-depth the third one in Section 2.2). The aim of this paper is to solve the fourth and last step: elucidate is the rate of decay to 0 of the *optimal* L^p-mean dual quantization error modulus, *i.e.* minimized over all grids Γ of size at most N – as N grows to infinity.

This is the to establish in a dual quantization framework the counterpart of Zador's Theorem which rules the convergence rate of optimal "regular" (Voronoi) quantization and is recalled below. To be more precise, we will establish such a theorem, for L^∞-bounded r.v.s but also, once mean dual quantization error will have been extended in an appropriate way following [12], to general r.v.s.

Let us now introduce in more formal way the (local and mean) dual quantization error moduli, following [12]. For a grid $\Gamma \subset \mathbb{R}^d$, we define the *$L^p$-mean dual quantization error* of X induced by the grid Γ by

$$d_p(X;\Gamma) = \|F_p(X;\Gamma)\|_{L^p(\mathbb{P})} \tag{5}$$

where F_p denotes the *local dual quantization error* function defined by

$$F_p(\xi;\Gamma) \quad = \quad \inf\left\{\Big(\sum_{x\in\Gamma}\lambda_x\|\xi - x\|^p\Big)^{\frac{1}{p}},\ \lambda_x \in [0,1],\ \sum_{x\in\Gamma}\lambda_x\, x = \xi,\ \sum_{x\in\Gamma}\lambda_x = 1\right\} \tag{6}$$

Note that $F_p(\xi;\Gamma) < +\infty$ if and only if $\xi \in \mathrm{conv}(\Gamma)$ so that $d_p(X;\Gamma) < +\infty$ if and only if $X \in \mathrm{conv}(\Gamma)$ $\mathbb{P}$-*a.s.*. and that $d_p(X;\Gamma) = \left\|X - \widehat{X}^{\Gamma,dual}\right\|^p_{L^p(\mathbb{P}\otimes\mathbb{P}_0)}$. Hence, this notion only makes sense for compactly supported r.v.s. In particular if the support of $\mathbb{P}_X$ is compact and contains $d+1$ affinely independent points, $d_{n,p}(X,\Gamma) = +\infty$ as long as $n \le d$.This new quantization modulus leads to an optimal dual quantization problem at *level* N,

$$d_{n,p}(X) = \inf\Big\{d_{n,p}(X,\Gamma),\ \Gamma \subset \mathbb{R}^d,\ |\Gamma| \le n\Big\} = \inf\Big\{\|F_p(X;\Gamma)\|_p,\ \Gamma \subset \mathbb{R}^d,\ |\Gamma| \le n\Big\}. \tag{7}$$

One important application of quantization in general is the use of quantization grids as numerical cubature formula (see (3)). The main feature here is the stationarity which allows to derive a second order formula for the integration error. Since, by construction, dual quantization can achieve stationarity only on a compact set, we show in section 2.2 that the extension of dual quantization to non-compactly supported random variables as defined in [12] preserves this second order rate on the whole support of the r.v.

We therefore define the splitting operator $\mathcal{J}_\Gamma$ outside $\mathrm{conv}(\Gamma)$ by setting

$$\forall\, \xi \in \mathbb{R}^d \setminus \mathrm{conv}(\Gamma), \quad \mathcal{J}_\Gamma(\omega_0,\xi) = \mathrm{Proj}_{\mathrm{conv}\,(\Gamma)\cap\partial\Gamma}(\xi)$$

where $\mathrm{Proj}_{\mathrm{conv}\,(\Gamma)\cap\partial\Gamma}$ is a Borel nearest neighbour projection on $\mathrm{conv}\,(\Gamma) \cap \partial\Gamma$. This choice is not unique: an alternative extension could be to set $\mathcal{J}_\Gamma(\omega_0,\xi) = \mathrm{Proj}_{\mathrm{conv}(\Gamma)}(\xi)$. But the above choice is tractable in terms of simulation and we will prove that it does not deteriorate the resulting mean error when $|\Gamma| \to +\infty$. Though the stationary property is lost as expected, we point out in Section 2.2 that this operator remains as performing as $\mathcal{J}_\Gamma$ is for bounded r.v.s when implementing cubature formulas for unbounded r.v.s.

Then, we to derive the *extended local dual quantization error* function by

$$\bar{F}_p(\xi;\Gamma) := F_p(\xi;\Gamma)\,\mathbf{1}_{\mathrm{conv}(\Gamma)}(\xi) + \mathrm{dist}(X,\Gamma)\,\mathbf{1}_{\mathrm{conv}(\Gamma)^c}(\xi), \tag{8}$$

and the *extended L^p-mean dual quantization error* of X induced by Γ by

$$\bar{d}_p(X;\Gamma) = \|\bar{F}_p(X;\Gamma)\|_{L^p(\mathbb{P})}. \tag{9}$$

Finally, we define the *extended L^p-mean dual quantization error* at level n given by

$$\bar{d}_{n,p}(X) = \inf\Big\{\bar{d}_p(X,\Gamma),\ \Gamma\subset\mathbb{R}^d,\ |\Gamma|\le n\Big\}. \tag{10}$$

Finally, we briefly recall a few facts about the (regular) *Voronoi optimal quantization problem* at level n associated to the nearest neighbour projection Proj_Γ: it reads

$$e_{n,p}(X) = \inf\left\{\|\mathrm{dist}(X,\Gamma)\|_{L^p(\mathbb{P})},\ \Gamma\subset\mathbb{R}^d,\ |\Gamma|\le n\right\} \tag{11}$$

(where $\mathrm{dist}(x,A) = \inf_{a\in A}\|x-a\|$). It is well-known that $e_{n,p}(X)\downarrow 0$ as soon as $n\to+\infty$ and $X\in L^p(\mathbb{P})$. Moreover, the rate of convergence to 0 of $e_{n,p}(X)$ is ruled by Zador's Theorem (see [5]).

Theorem 1 (Zador)**.** *Let $X\in L^{p'}_{\mathbb{R}^d}(\mathbb{P})$, $p'>p$. Let $\mathbb{P}_X = h.\lambda_d+\nu$, $\nu\perp\lambda_d$ be the distribution of X where λ_d denotes the Lebesgue measure on $(\mathbb{R}^d,\mathcal{B}or(\mathbb{R}^d))$. Then*

$$\lim_{n\to\infty} n^{\frac{1}{d}} e_{n,p}(X) = Q^{vq}_{\|\cdot\|,p,d}\,\|h\|^{\frac{1}{p}}_{\frac{d}{p+d}}$$

where $\|h\|_{\frac{d}{p+d}} = \Big(\displaystyle\int_{\mathbb{R}^d} h(\xi)^{\frac{d}{p+d}}d\xi\Big)^{1+\frac{p}{d}}$ *and* $Q^{vq}_{\|\cdot\|,p,d} = \inf_n n^{\frac{1}{d}}\, e_{n,p}(U([0,1]^d))\in(0,\infty)$.

The above rate depends on d and is known as the *curse of dimensionality.* Its statement and proof goes back to Zador (PhD, 1954) for the uniform distributions on hypercubes, its extension to absolutely continuous distributions is due to Bucklew and Wise in [2]. A first general rigor proof (according to mathematical standards) was provided in [5] in 2000 (see also [6] for a survey of the history of quantization).

It should be noted that $d_{n,p}(X)$ and $\bar{d}_{n,p}(X)$ do not coincide even for bounded r.v.s. We will extensively use (see [12]) that

$$d_{n,p}(X)\ge\bar{d}_{n,p}(X)\ge e_{n,p}(X).$$

This paper is entirely devoted to establishing the sharp asymptotics of the optimal dual quantization error moduli $d_{n,p}(X)$ and $\bar{d}_{n,p}(X)$ as n goes to infinity. The main result is stated in Theorem 2 (Zador's like theorem) (see Section 2.1 below). Proposition 2 (a Pierce like Lemma) is a companion result which provides a non-asymptotic upper bound for the exact rate simply involving moments of the r.v. X (higher than p). Our proof has the same structure as that of the original Zador Theorem (see *e.g.* [5] where it has been rigorously completed for the first time), except that the splitting operator $\mathbb{J}_\Gamma$ is much more demanding to handle than the plain nearest neighbour projection: it requires more sophisticate arguments borrowed from convex analysis (including dual primal/methods) and geometry, both in a probabilistic framework. In one dimension the exact rate $O(n^{-1})$ for $d_{n,p}(X)$ and $\bar{d}_{n,p}(X)$ follows from a random quantization argument detailed in Section 4 (extended Pierce Lemma for $d_{n,p}(X)$). This rate can be transferred in a d-dimensional framework to $O(n^{-\frac{1}{d}})$ using a product (dual) quantization argument (see Proposition 1 below and Section 3.2). Finally, the sharp upper bound is obtained in Section 5 by successive approximation procedures of the density of X, whereas the lower bound relies on a new "firewall" Lemma.

NOTATIONS: • $\operatorname{conv}(A)$ stands for the convex hull of $A \subset \mathbb{R}^d$, $|A|$ for its cardinality, $\operatorname{diam}_{\|.\|}(A) = \sup_{x,y\in A}\|x-y\|$ for its diameter and $\operatorname{aff.dim}(A)$ for the dimension of the affine subspace of $\mathbb{R}^d$ spanned by A.
• We denote $\binom{n}{i} := \frac{n!}{i!(n-i)!}$, n, $i \in \{0,\ldots,n\}$, $n \in \mathbb{N}$.
• $\lfloor x \rfloor$ and $\lceil x \rceil$ will denote the lower and the upper integral part of the real number x respectively; set likewise $x_\pm = \max(\pm x, 0)$. For two sequences of real numbers (a_n) and (b_n), $a_n \sim b_n$ if $a_n = u_n b_n$ with $\lim_n u_n = 1$.
• For every $x = (x^1,\ldots,x^d) \in \mathbb{R}^d$, $|x|_{\ell^r} = (|x^1|^r + \cdots |x^d|^r)^{1/r}$ denotes the ℓ^r-norm or pseudo-norm, $0 < r < +\infty$ and $|x|_{\ell^\infty} = \max_{1\le i\le d}|x_i|$ denotes the ℓ^∞-norm. A general norm on $\mathbb{R}^d$ will be denoted $\|\cdot\|$.
• $\operatorname{supp}(\mu)$ denotes the support of a distribution μ on $(\mathbb{R}^d, \mathcal{B}or(\mathbb{R}^d))$.

2 Main results and motivation for extended dual quantization

2.1 Main results

The theorem below establishes for any $p > 0$ and any norm on $\mathbb{R}^d$ the counterpart of Zador's Theorem in the framework of dual quantization for both $d_{n,p}$ and $\bar{d}_{n,p}$ error moduli.

Theorem 2. *(a) Let $X \in L^\infty_{\mathbb{R}^d}(\mathbb{P})$. Assume the distribution $\mathbb{P}_X$ of X reads $\mathbb{P}_X = h.\lambda_d + \nu$, $\nu \perp \lambda_d$. Then*

$$\lim_{n\to\infty} n^{\frac{1}{d}}\, d_{n,p}(X) = \lim_{n\to\infty} n^{\frac{1}{d}}\, \bar{d}_{n,p}(X) = Q^{dq}_{\|\cdot\|,p,d}\, \|h\|^{\frac{1}{p}}_{\frac{d}{p+d}}$$

where $Q^{dq}_{\|\cdot\|,p,d} = \inf_{n\ge1} n^{\frac{1}{d}}\, d_{n,p}(U([0,1]^d)) \in (0,\infty)$.

(b) Let $X \in L^{p'}_{\mathbb{R}^d}(\mathbb{P})$, $p' > p$. Assume the distribution $\mathbb{P}_X$ of X reads $\mathbb{P}_X = h.\lambda_d + \nu$, $\nu \perp \lambda_d$. Then

$$\lim_{n\to\infty} n^{\frac{1}{d}}\, \bar{d}_{n,p}(X) = Q^{dq}_{\|\cdot\|,p,d}\, \|h\|^{\frac{1}{p}}_{\frac{d}{p+d}}.$$

(c) If $d = 1$, then

$$d_{n,p}(U([0,1])) = \left(\frac{2}{(p+1)(p+2)}\right)^{\frac{1}{p}} \frac{1}{n-1},$$

which implies $Q^{dq}_{|\,.\,|,p,1} = \left(\frac{2^{p+1}}{p+2}\right)^{\frac{1}{p}} Q^{vq}_{|\,.\,|,p,1}$.

Moreover, we will also establish in Section 5 an upper bound for the dual quantization coefficient $Q^{\mathrm{dq}}_{\|\cdot\|,p,d}$ when $\|\cdot\| = |\cdot|_{\ell^r}$.

Proposition 1 (Product quantization)**.** *Let $r, p \in [1,\infty)$ with $r \le p$. Then it holds for every $d \in \mathbb{N}$*

$$Q^{dq}_{|\cdot|_{\ell^r},p,d} \le d^{\frac{1}{r}} \cdot Q^{dq}_{|\cdot|,p,1}$$

where $|\,.\,|$ denotes standard absolute value on $\mathbb{R}$.

Since this upper bound achieves the same asymptotic rate as in the case of regular quantization (cf. Corollary 9.4 in [5]), this suggests the rate $O(d^{\frac{1}{r}})$ to be also the true one for $Q^{\rm dq}_{\|\cdot\|,p,d}$ as $d\to\infty$.

As a step towards the above sharp rate theorem, we also establish a counterpart of the so-called Pierce Lemma (as stated in an operating form *e.g.* in [7]). In practice, it turns out to be quite useful for applications since it provides non-asymptotic error bounds which only depend on the moments of the r.v. X and the size of the optimal grid as emphasized in [10] (see section 4.1 for the proof).

Proposition 2 (d-dimensional extended Pierce Lemma)**.** (a) *Let* $p,\,\eta>0$. *There exists a real constant* $C_{d,p,\eta}>0$ *such that, for every* $n\geq 1$ *and every r.v.* $X\in L^{p+\eta}_{\mathbb{R}^d}(\Omega,\mathcal{A},\mathbb{P})$,

$$\bar{d}_{n,p}(X)\leq C_{d,p,\eta}\sigma_{p+\eta,\|\cdot\|}(X)\,n^{-1/d}$$

where $\sigma_{p+\eta,\|\cdot\|}(X)=\inf_{a\in\mathbb{R}^d}\|X-a\|_{L^{p+\eta}}$ *denotes the* $L^{p+\eta}$*-pseudo-standard deviation of* X.
(b) *If* $\mathrm{supp}(\mathbb{P}_X)$ *is compact then there exists a real constant* $C'_{d,p,\eta}>0$ *such that, for every* $n\geq 1$

$$d_{n,p}(X)\leq C'_{d,p,\eta}\mathrm{diam}_{\|.\|}(\mathrm{supp}(\mathbb{P}_X))\,n^{-1/d}.$$

2.2 How to use the extended L^p-dual quantization error modulus?

We briefly explain why the extended dual quantization error modulus, already been introduced in [12] for non-compactly supported distributions, is the right tool to *perform automatically an optimized truncation* of non-compactly supported distributions. basically, it uses its additional "outer Voronoi projection" as a *penalization term* which expands automatically the convex hull of the dually optimal grid at its appropriate "amplitude", making altogether the distribution outside of its convex hull "negligible" and sharing an optimal rate of decay $n^{-\frac{1}{d}}$ as its size n goes to infinity. The specific choice of a Voronoi quantization among other possible solutions for this penalization is motivated by both its theoretical tractability and its simple implementability in stochastic grid optimization algorithms. This feature if of the highest importance for numerical integration or conditional execration approximation. This is the main motivation to introduce and deeply investigate the sharp asymptotics of this L^p-mean extended dual quantization error modulus $\bar{d}_{n,p}(X)$.

We saw in [12] that *Euclidean dual quantization* of a compactly supported distribution produces *stationary* (dual) quantizers, namely r.v.s $\widehat{X}^{dual}$ satisfying $\mathbb{E}(\widehat{X}^{dual}\,|\,X)=X$, so that (see Proposition ?? in [12]), dual quantization based cubature formula induce on functions $f\in\mathcal{C}^1_{\rm Lip}(\mathbb{R}^d,\mathbb{R})$ (Lipschitz functions with Lipschitz continuous gradient) an error at most equal to $[Df]_{\rm Lip}d_{2,n}(X)^2$. Taking into account the rate established in Theorem 2(a), this yields a $O(n^{-\frac{2}{d}})$ error rate.

There is no way to extend dual quantization to (possibly) unbounded r.v.s so that it preserves the above stationarity property. However, with the choice we made (nearest neighbor projection on the grid outside its convex hull), natural heuristic arguments strongly suggest that the above order $O(n^{-\frac{2}{d}})$ is still satisfied for functions in $\mathcal{C}^1_{\rm Lip}(\mathbb{R}^d,\mathbb{R})$.

We consider an unbounded Borel distribution $\mu=\mathbf{P}_X$ of an $\mathbb{R}^d$-valued r.v. X. Let Γ_n be an *Euclidean* L^2*-optimal extended* dual quantization grid of size n for μ (see [12] or Theorem 4) and $\widehat{X}^{dual}$ the resulting Γ_n-valued extended dual quantization of X. Let $C_n=\mathrm{conv}\big(\Gamma_n\big)$ denote the convex hull of Γ_n. It is clear by construction of $\widehat{X}^{dual}$ that $\widehat{X}^{dual}=\widetilde{X}^{dual}+\widetilde{X}^{vor}$ where, with obvious notations,

$$\mathbf{1}_{\{X\in C_n\}}\mathbb{E}\big(\widetilde{X}^{dual}|X\big)=\mathbf{1}_{\{X\in C_n\}}X\ \ \text{(dual stationarity) and}\ \ \widetilde{X}^{vor}=\mathrm{Proj}_{\Gamma_n\cap C_n}(X).$$

Hence, if $f \in \mathcal{C}^1_{\rm Lip}(\mathbb{R}^d,\mathbb{R})$, $\mathbb{E}\big((Df(X)|X-\widetilde{X}^{dual})|X\in C_n\big)=0$ and

$$\begin{aligned}\Big|\mathbb{E}\Big(f(\widetilde{X}^{dual})|X\in C_n\Big)-\mathbb{E}\Big(f(X)|X\in C_n\Big)\Big| &= \Big|\mathbb{E}\left(f(\widetilde{X}^{dual})-f(X)-Df(X).(X-\widetilde{X}^{dual})|X\in C_n\right)\Big|\\ &\le [Df]_{\rm Lip} d_{n,2}(\Gamma_n,\widetilde{X}^{dual}|X\in C_n)^2.\end{aligned}$$

Consequently,

$$\begin{aligned}\Big|\mathbb{E}\Big(f(\widetilde{X}^{dual})\mathbf{1}_{\{X\in C_n\}}\Big)-\mathbb{E}\Big(f(X)\mathbf{1}_{\{X\in C_n\}}\Big)\Big| &\le [Df]_{\rm Lip} d_{n,2}(\widetilde{X}^{dual},\Gamma_n)^2/\mathbf{P}(X\in C_n)\\ &\le [Df]_{\rm Lip} \bar{d}_{n,2}(X,\Gamma_n)^2/\mathbf{P}(X\in C_n).\end{aligned}$$

On the other hand,

$$\begin{aligned}\Big|\mathbb{E}\Big(f(\widetilde{X}^{vor})\mathbf{1}_{\{X\notin C_n\}}\Big)-\mathbb{E}\Big(f(X)\mathbf{1}_{\{X\notin C_n\}}\Big)\Big| &\le [f]_{\rm Lip} e_{n,2}(X,\Gamma_n)\mathbf{P}\big(X\notin C_n\big)^{\frac12}\\ &\le [f]_{\rm Lip} \bar{d}_{n,2}(X)\mathbf{P}\big(X\notin C_n\big)^{\frac12}.\end{aligned}$$

Relying on Theorem 2(b), we know that, if $\mu = h.\lambda_d \overset{\perp}{+} \nu$, then $\bar{d}_{n,2}(X)\sim Q^{dq}_{2,|\cdot|_{eucl}}\|h\|^{\frac1p}_{\frac{d}{2+d}} n^{-\frac1d}$. The "outside" contribution will be negligible compared to the "inside" one as soon as

$$\mathbf{P}\big(X\notin C_n\big) = o\Big(\bar{d}_{n,2}(X,\Gamma_n)^2\Big) = o\Big(n^{-\frac2d}\Big). \tag{12}$$

This condition turns out to be not very demanding and can be checked, at least heuristically, as illustrated below: if $X \overset{d}{=} \mathcal{N}(0;I_d)$, one may conjecture, taking advantage of the spherical symmetries of the normal distribution, that C_n is approximately a sphere centered at 0 with radius $\rho_n = \max_{a\in\Gamma_n}|a|$. As

$$\mathbf{P}(|X|\ge\xi)\sim V_d\,\xi^{d-2}e^{-\frac{\xi^2}{2}} \;\text{ as }\; \xi\to+\infty \quad (\text{with } V_d=\lambda_{d-1}(S_d(0,1))).$$

Condition (12) is satisfied as soon as $\liminf_n \frac{\rho_n}{\sqrt{\log n}} > \frac{2}{\sqrt{d}}$ ($\ge$ if $d=1,2$). As an example, one must have in mind that, for optimal *Voronoi* quantization, this inequality is satisfied since (see [9]) $\lim_n \frac{\rho_n}{\sqrt{\log n}} = \sqrt{2(1+2/d)} > \frac{2}{\sqrt d}$. More precisely, we have

$$\mathbf{P}\big(X\notin C_n\big)\sim\kappa_d(\log n)^{\frac d2-1}n^{-1-\frac2d} \;\text{ so that }\; \bar{d}_{n,2}(X)\mathbf{P}\big(X\notin C_n\big)^{\frac12} = O\big(n^{-\frac2d-\frac12}(\log n)^{\frac{d-2}{4}}\big).$$

Numerical experiments, not reproduced here, carried out with the above $\mathcal{N}(0;I_d)$ distribution confirm that the radius of optimal dual quantizers always achieves this asymptotics which makes the above partially heuristic reasoning very likely. Moreover, we also tested the two rates of convergence of $\mathbb{P}(X\in C_n)$ and $\bar{d}_{n,2}(X)^2$, this time on the joint distribution of the $(W_1,\sup_{t\in[0,1]}W_t)$, W standard Brownian motion which has less symmetries (see appendix A). They also confirm that the above partially heuristic reasoning is very likely.

3 Dual quantization: background and basic properties

Throughout the paper, except specific mention, $\mathbb{R}^d$ is equipped with a norm $\|\cdot\|$.

3.1 More background

In the introduction, the definitions related to Voronoi (or regular)and dual quantizations of a r.v. X defined on a probability space $(\Omega,\mathcal{S},\mathbb{P})$ have been recalled (see (7)-(10)). The aim of this section is to come back briefly to the origin and the motivations which led us to introduce dual quantization in [12]. On the way, we will also recall several basic results on dual quantization established in [12]. First, we will assume throughout the paper that the r.v. of interest, X, is truly d-dimensional in the sense that

$$\text{aff.dim}(\text{supp}(\mathbb{P}_X)) = d.$$

Let us start by a few practical points. First note that although all these definitions are related to a r.v. X, in fact it only depends on the distribution $\mathbf{P} = \mathbb{P}_X$, so we will also often write $d_p(\mathbf{P}, \Gamma)$ for $d_p(X, \Gamma)$ and $d_{n,p}(\mathbf{P})$. Furthermore, to alleviate notations, we will denote from now on F^p, d^p and $\bar{d}^p$, ... instead of $(F_p)^p$, $(d_p)^p$ and $(\bar{d}_p)^p$,...

Let us come back to the terminology *dual quantization*: it refers to a canonical example of the intrinsic stationary splitting operator: the dual quantization operator.

To be more precise, let $p \in [1, +\infty)$ and let $\Gamma = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ be a grid of size $n \geq d+1$ such that $\text{aff.dim}(\Gamma) = d$ *i.e.* Γ contains at least one $d+1$-tuple of affinely independent points.

The underlying idea is to "split" $\xi \in \text{conv}(\Gamma)$ across at most $d+1$ affinely independent points in Γ proportionally to its barycentric coordinates of ξ. There are usually many possible choices of such a Γ-valued $(d+1)$-tuple of affinely independent points, so we introduced a minimal inertia based criterion to select the most appropriate one ξ, namely the function $F_p(\xi; \Gamma)$ defined for every ξ as the value of the minimization problem

$$F_p(\xi; \Gamma) = \inf_{(\lambda_1, \ldots, \lambda_n)} \Big\{ \Big(\sum_{i=1}^{n} \lambda_i \|\xi - x_i\|^p \Big)^{\frac{1}{p}}, \lambda_i \in [0,1], \sum_i \lambda_i \begin{bmatrix} x_i \\ 1 \end{bmatrix} = \begin{bmatrix} \xi \\ 1 \end{bmatrix} \Big\}. \tag{13}$$

Owing to the compactness of the constraint set ($\lambda_i \geq 0$, $\sum_i \lambda_i = 1$, $\sum_i \lambda_i x_i = \xi$), there exists at least one solution $\lambda^*(\xi)$ to the above minimization problem. Moreover, for any such solution, one shows using convex extremality arguments, that the set $I^*(\xi) := \big\{ i \in \{1, \ldots, n\} \text{ s.t. } \lambda_i^*(\xi) > 0 \big\}$ defines an affinely independent subset $\{x_i,\ i \in I^*(\xi)\}$.

If, for every $\xi \in conv(\Gamma)$, this solution is unique, the *dual quantization operator* is simply defined on $\text{conv}(\Gamma)$ by

$$\forall\, \xi \in \text{conv}(\Gamma),\ \forall\, \omega_0 \in \Omega_0, \quad \mathcal{J}^*_\Gamma(\omega_0, \xi) = \sum_{i \in I(\xi)^*} x_i \mathbf{1}_{\{\sum_{j=1}^{i-1} \lambda_j^*(\xi) \leq U(\omega_0) < \sum_{j=1}^{i} \lambda_j^*(\xi)\}}, \tag{14}$$

where U denotes a random variable uniformly distributed over $[0,1]$ on an exogenous probability space $(\Omega_0, \mathcal{S}_0, \mathbb{P}_0)$. This operator $\mathcal{J}^*_\Gamma$ is then measurable (see [12]).

The above uniqueness assumption is not so stringent, especially for applications. Thus, in a purely Euclidean quadratic framework: $\|\,.\,\| = |\,.\,|_{\ell^2}$ (canonical Euclidean norm) and $p = 2$ and if Γ is said in "general position" ([1]), then $\Big\{ \{\xi \text{ s.t. } I^*(\xi) = I\},\ |I| \leq d+1 \Big\}$ makes up a Borel partition of $\text{conv}(\Gamma)$ (with possibly empty elements), known in 2-dimension as the *Delaunay triangulation* of Γ (see [14] for the connection with Delaunay triangulations).

In a more general framework, we refer to [12] for a construction of dual quantization operators. Such operators are splitting operators since, by construction, they satisfy the stationarity property (1).

One must have in mind that the dual quantization operators $\mathcal{J}^*_\Gamma(\omega_0, \xi)$ play the role of the nearest neighbour projections for regular Voronoi quantization. One checks that, by construction,

$$\forall\, \xi \in \text{conv}(\Gamma), \quad \|\mathcal{J}^*_\Gamma(\xi) - \xi\|_{L^p(\mathbb{P}_0)} = \|F_p(\xi; \Gamma)\|_{L^p(\mathbb{P}_0)}$$

[1] no $d+2$ points of Γ lie on a sphere in $\mathbb{R}^d$.

so that, as soon as $\mathrm{supp}(\mathbb{P}_X) \subset \Gamma$ (or equivalently $\mathbb{P}(X \in \mathrm{conv}(\Gamma)) = 1$),

$$d_p(X;\Gamma) = \|\mathcal{J}^*_\Gamma(X) - X\|_{L^p(\mathbb{P}_0\otimes\mathbb{P})} = \|F_p(X;\Gamma)\|_{L^p(\mathbb{P}_0\otimes\mathbb{P})}.$$

At this stage, it appears naturally that the the second step of the optimization process is to find (at least) one grid which optimally "fits" (the distribution of) X for this criterion *i.e.* which is the solution to the second level optimization problem

$$d_{n,p}(X) = \inf\big\{\|\mathcal{J}^*_\Gamma(X) - X\|_{L^p(\mathbb{P}_0\otimes\mathbb{P})},\ \mathcal{J}^*_\Gamma : \Omega_0 \times \mathrm{conv}(\Gamma) \to \Gamma, \mathrm{conv}(\Gamma) \supset \mathrm{supp}(\mathbb{P}_X),\ |\Gamma| \le n\big\}.$$

Note that if $X \in L^\infty_{\mathbb{R}^d}(\mathbb{P})$, $d_{n,p}(X) < +\infty$ if and only if $n \ge d+1$ (whereas it is identically infinite if X is not essentially bounded). The existence of an optimal grid (or dual quantizer) has been established in [12] (see below).

The error modulus $d_{n,p}(X)$ can also be characterized as the *lowest L^p-mean approximation error by a r.v. having at most n values and satisfying the intrinsic stationarity property* as established in [12] (Theorem 2, precisely recalled in Theorem 3 below). It should be compared to the well-known property satisfied by the mean (regular) quantization error modulus $e_{n,p}(X)$, namely

$$e_{n,p}(X) = \inf\Big\{\|X - \widehat{X}\|_{L^p(\mathbb{P})},\ |\widehat{X}(\Omega)| \le n\Big\}.$$

A stochastic optimization procedure based on a stochastic gradient approach has been devised in [12] to compute optimal dual quantization grids w.r.t. various distributions (so far, uniform over $[0,1]^2$, normal, $(W_1, \sup_{t\in[0,1]} W_t)$, W standard Brownian motion in a purely Euclidean framework).

Let us conclude by two results established in [12]. The first one is the characterization of dual quantization operator in terms in terms of best L^p-approximation (see [12], Theorem 2).

Theorem 3. *Let $X : \Omega, \mathcal{S}, \mathbb{P}) \to \mathbb{R}^d$ be a r.v. such that* $\mathrm{aff.dim}(\mathrm{supp}(\mathbb{P}_X)) = d$ *and let $n \in \mathbb{N}$, $n \ge d+1$. Then*

$$\begin{aligned} d_{n,p}(X) &= \inf\big\{\mathbb{E}\|X - \mathcal{J}_\Gamma(X)\|_{L^p} : \mathcal{J}_\Gamma : \Omega_0 \times \mathbb{R}^d \to \Gamma,\ \textit{intrinsic stationary}, \\ &\qquad\qquad \mathrm{supp}(\mathbb{P}_X) \subset \mathrm{conv}(\Gamma),\ |\Gamma| \le n\big\} \\ &= \inf\big\{\mathbb{E}\|X - \widehat{X}\|_{L^p} : \widehat{X} : (\Omega_0 \times \Omega, \mathcal{S}_0 \otimes \mathcal{S}, \mathbb{P}_0 \otimes \mathbb{P}) \to \mathbb{R}^d, \\ &\qquad\qquad |\widehat{X}(\Omega_0 \times \Omega)| \le n,\ \mathbb{E}(\widehat{X}|X) = X\big\} \le +\infty. \end{aligned}$$

This quantity is finite if and only if $X \in L^\infty(\Omega, \mathcal{S}, \mathbb{P})$.

Finally, the following existence result for optimal dual quantizers *at level $n \in \mathbb{N}$* and the L^p-norm with $p \in (1,\infty)$ is established in [12]. Although we will not use it in our proofs, this result is recalled for the reader's convenience.

Theorem 4 (Existence of optimal quantizers)**.** *Let $X \in L^p(\mathbb{P})$ for some $p \in (1,\infty)$.*

(a) If $\mathrm{supp}(\mathbb{P}_X)$ is compact, then there exists for every $n \in \mathbb{N}$ a grid $\Gamma^*_n \subset \mathbb{R}^d$, $|\Gamma^*_n| \le n$ such that $d_p(X;\Gamma^*_n) = d_{n,p}(X)$.

(b) If $\mathbb{P}_X$ is strongly continuous in the sense that it assigns no mass to hyperplanes of $\mathbb{R}^d$, then there exists for every $n \in \mathbb{N}$ a grid $\Gamma^*_n \subset \mathbb{R}^d$, $|\Gamma^*_n| \le n$ such that $\bar{d}_p(X;\Gamma^*_n) = \bar{d}_{n,p}(X)$.

If furthermore $|\mathrm{supp}(\mathbb{P}_X)| \ge n$, then the above statements hold with $|\Gamma^*_n| = n$.

3.2 Local properties of the dual quantization functional

We establish or recall in this paragraph some first general properties of the local L^p-dual quantization functional F^p, which will be needed for the final proof of Theorem 2.

Proposition 3. *Let Γ_1, $\Gamma_2 \subset \mathbb{R}^d$ be finite grids and let $\xi \in \mathbb{R}^d$. Then*

$$\Gamma_1 \subset \Gamma_2 \Longrightarrow F_p(\xi;\Gamma_2) \le F_p(\xi;\Gamma_1).$$

Proof. First note that the set $\{\lambda \in \mathbb{R}^n \mid \left[\begin{smallmatrix} x_1 & \cdots & x_m \\ 1 & \cdots & 1 \end{smallmatrix}\right] \lambda = \left[\begin{smallmatrix} \xi \\ 1 \end{smallmatrix}\right]\}$ is clearly a compact set on which the continuous function $\lambda \mapsto \sum_{i=1}^n \lambda_i \|\xi - x_i\|^p$ attains a minimum. Assume $\Gamma_1 = \{x_1, \ldots, x_m\}$ and $\Gamma_2 = \{x_1, \ldots, x_m, x_{m+1}, \ldots, x_n\}$. Then

$$\begin{aligned} F^p(\xi;\Gamma_2) = & \min_{\substack{\lambda \in \mathbb{R}^n \\ \text{s.t. } \left[\begin{smallmatrix} x_1 & \cdots & x_n \\ 1 & \cdots & 1 \end{smallmatrix}\right]\lambda = \left[\begin{smallmatrix} \xi \\ 1 \end{smallmatrix}\right],\, \lambda \ge 0}} \sum_{i=1}^n \lambda_i \,\|\xi - x_i\|^p \le \min_{\substack{\lambda \in \mathbb{R}^n, \lambda_{m+1} = \cdots = \lambda_n = 0 \\ \text{s.t. } \left[\begin{smallmatrix} x_1 & \cdots & x_m \\ 1 & \cdots & 1 \end{smallmatrix}\right]\lambda = \left[\begin{smallmatrix} \xi \\ 1 \end{smallmatrix}\right],\, \lambda \ge 0}} \sum_{i=1}^n \lambda_i \,\|\xi - x_i\|^p \\ & = \min_{\substack{\lambda \in \mathbb{R}^m \\ \text{s.t. } \left[\begin{smallmatrix} x_1 & \cdots & x_m \\ 1 & \cdots & 1 \end{smallmatrix}\right]\lambda = \left[\begin{smallmatrix} \xi \\ 1 \end{smallmatrix}\right],\, \lambda \ge 0}} \sum_{i=1}^m \lambda_i \,\|\xi - x_i\|^p = F^p(\xi;\Gamma_1). \qquad \square \end{aligned}$$

We will also make use of the following three properties established in [12] (Propositions 11, 12, 13 respectively). In particular, the third claim yields a first upper bound for the asymptotics of the local L^p-dual quantization error when the size of the grid goes to infinity.

Proposition 4. (a) Scalar bound: *Let $\Gamma = \{x_1, \ldots, x_n\} \subset \mathbb{R}$ with $x_1 \le \ldots \le x_n$. Then*

$$\forall \xi \in [x_1, x_n], \quad F^p(\xi;\Gamma) \le \max_{1 \le i \le n-1} \Big(\frac{x_{i+1} - x_i}{2}\Big)^p.$$

(b) Local product Quantization: *Let $\|\cdot\| = |\cdot|_{\ell^p}$ and let $\Gamma = \prod_{1 \le j \le d} \Gamma_j$ for some $\Gamma_j \subset \mathbb{R}$. Then*

$$\forall \xi \in \mathbb{R}^d, \quad F_{p,|\cdot|_{\ell^p}}(\xi;\Gamma) = \Big(\sum_{j=1}^d F^p(\xi^j;\Gamma_j)\Big)^{\frac{1}{p}}$$

and the same holds true with $\bar{F}_{p,\ell^p}$ on $\mathbb{R}^d$.

(c) Product Quantization: *Let $C = a + L\,[0,1]^d$, $a = (a_1, \ldots, a_d) \in \mathbb{R}^d$, $L > 0$, be a hypercube, with edges parallel to the coordinate axis with common edge-length L. Let Γ be the product quantizer of size $(m+1)^d$ defined by*

$$\Gamma = \prod_{k=1}^d \Big\{a_j + \frac{iL}{m},\, i = 0, \ldots, m\Big\}.$$

There exists a positive real constant $C_{\|\cdot\|,p} = \sup_{|x|_{\ell^p}=1} \|x\|^p > 0$ such that

$$\forall \xi \in C, \quad F^p(\xi;\Gamma) \le d\, C_{\|\cdot\|,p} \cdot \Big(\frac{L}{2}\Big)^p \cdot m^{-p}. \tag{15}$$

4 Extended Pierce lemma and applications

The aim of this section is to provide a non-asymptotic "universal" upper-bound for the optimal (extended) L^p-mean dual quantization error in the spirit of [13]: it achieves nevertheless the optimal rate of convergence when the size n goes to infinity. Like for Voronoi quantization this upper-bound deeply relies on a random quantization argument and will be a key in the proof of the sharp rate (step 2 of the proof of Theorem 2).

For every integer $n \geq 1$, we define the set of "non-decreasing" n-tuples of $\mathbb{R}^n$ by

$$\mathcal{I}_n := \{(x_1, \ldots, x_n) \in \mathbb{R}^n,\ -\infty < x_1 \leq x_2 \leq \cdots \leq x_n < +\infty\}.$$

Let $(x_1, \ldots, x_n) \in \mathcal{I}_n$ (so that $\Gamma = \{x_1, \ldots, x_n\}$ has at most n elements) and let $\xi \in \mathbb{R}$. When $d = 1$, it is clear that the minimization problem (6) always has a unique solution when $\xi \in [x_1, x_n]$ so that, for every $\omega_0 \in \Omega_0 = [0, 1]$, one has

$$\begin{aligned}\bar{\mathcal{J}}^*_{(x_1,\ldots,x_n)}(\omega_0, \xi) \quad &= \quad \sum_{i=1}^{n-1} \Big(x_i \mathbf{1}_{\{\omega_0 \leq \frac{x_{i+1}-\xi}{x_{i+1}-x_i}\}} + x_{i+1} \mathbf{1}_{\{\omega_0 \geq \frac{x_{i+1}-\xi}{x_{i+1}-x_i}\}} \Big) \mathbf{1}_{[x_i, x_{i+1})}(\xi) \\ &\quad + x_1 \mathbf{1}_{(-\infty, x_1)}(\xi) + x_n \mathbf{1}_{[x_n, +\infty)}(\xi).\end{aligned}$$

It follows from (8) that

$$\begin{aligned}\bar{F}^p_n(\xi, x_1, \ldots, x_n) \quad &= \quad \mathbb{E}_{\mathbb{P}_0} \big|\xi - \bar{\mathcal{J}}^*_{(x_1,\ldots,x_n)}(\omega_0, \xi)\big|^p \\ &= \quad \sum_{i=1}^{n-1} \left(\frac{(x_{i+1} - \xi)^p (\xi - x_i)}{x_{i+1} - x_i} + \frac{(x_{i+1} - \xi)(\xi - x_i)^p}{x_{i+1} - x_i} \right) \mathbf{1}_{[x_i, x_{i+1})}(\xi) \\ &\quad + (x_1 - \xi)^p \mathbf{1}_{(-\infty, x_1)}(\xi) + (\xi - x_n)^p \mathbf{1}_{[x_n, +\infty)}(\xi)\end{aligned} \tag{16}$$

(the subscript $_n$ is temporarily added to the functional $\bar{F}^p$, $\bar{F}_p$, etc, to emphasize that they are defined on $\mathcal{I}_n \times \mathbb{R}$). The functionals $\bar{F}^p_n$ share three important properties extensively used in what follows:

- *Additivity*: Let $(x_1, \ldots, x_{i_0}, \ldots, x_n) \in \mathcal{I}_n$. Then for every $\xi \in \mathbb{R}$

$$\bar{F}^p_n(\xi, x_1, \ldots, x_n) = \bar{F}^p_{i_0}(\xi, x_1, \ldots, x_{i_0}) \mathbf{1}_{(-\infty, x_{i_0})}(\xi) + \bar{F}^p_{n-i_0+1}(\xi, x_{i_0}, \ldots, x_n) \mathbf{1}_{[x_{i_0}, +\infty)}(\xi).$$

- *Consistency and monotony*: Let $(x_1, \ldots, x_n) \in \mathcal{I}_n$ and $\widetilde{x}_i \in [x_i, x_{i+1}]$ for an $i \in \{1, \ldots, n-1\}$. For every $\xi \in \mathbb{R}$,

$$\bar{F}^p_{n+1}(\xi, x_1, \ldots, x_{i-1}, x_i, \widetilde{x}_i, x_{i+1}, \ldots, x_n) \leq \bar{F}^p_n(\xi, x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n).$$

  When $\xi \in [x_1, x_n]$, $\bar{F}^p_n(\xi; x_1, \ldots, x_n)$ coincides with $F^p(\xi, \{x_1, \ldots, x_n\})$ and this inequality is a consequence of the definition of F_p as the value function of the minimization problem (6). Outside, the above inequality holds as an equality since it amounts to the nearest distance of ξ to $[x_1, x_n]$. As a consequence,

$$n \longmapsto \bar{d}_{n,p}(X) = \inf_{(x_1,\ldots,x_n) \in \mathcal{I}_n} \|\bar{F}_{p,n}(X, x_1, \ldots, x_n)\|_{L^p} \ \text{ is non-increasing,} \tag{17}$$

  More generally, for every fixed $x^0 \in \mathbb{R}$, both

$$n \longmapsto \inf_{(x^0, x_2, \ldots, x_n) \in \mathcal{I}_n} \|\bar{F}_{p,n}(X, x^0, x_2, \ldots, x_n)\|_{L^p} \ \text{ and } \ n \longmapsto \inf_{(x_1, x_2, \ldots, x_{n-1}, x^0) \in \mathcal{I}_n} \|\bar{F}_{p,n}(X, x_1, \ldots, x_{n-1}, x^0)\|_{L^p} \tag{18}$$

  are non-increasing.

- *Scaling*: $\forall\, \omega \in \Omega_0$, $\forall\, (x_1, \ldots, x_n) \in \mathcal{I}_n$, $\forall\, \xi \in \mathbb{R}$, $\forall\, \alpha \in \mathbb{R}_+$, $\forall\, \beta \in \mathbb{R}$,

$$\begin{aligned}(i) \quad \bar{F}^p_n(\alpha\, \xi + \beta, \alpha\, x_1 + \beta, \ldots, \alpha\, x_n + \beta) \quad &= \alpha\, \bar{F}^p_n(\xi, x_1, \ldots, x_n), \\ (ii) \qquad \bar{F}^p_n(\xi, x_1, \ldots, x_n) \quad &= \bar{F}^p_n(-\xi, -x_n, \ldots, -x_1).\end{aligned}$$

Theorem 5. *Let $p, \eta > 0$. There exists a real constant $C_{p,\eta} > 0$ such that for every random variable $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$,*

$$\forall\, n \geq 1, \qquad \inf_{(x_1,\ldots,x_n)\in\mathcal{I}_n} \|\bar{F}_{p,n}(X, x_1, \ldots, x_n)\|_{L^p} \leq C_{p,\eta}\|X\|_{L^{p+\eta}} n^{-1}.$$

The proof below relies on a random quantization argument involving an n-sample of the Pareto(δ)-distribution on $[1, +\infty)$. Though significantly more demanding, it plays the same crucial role in establishing the sharp rate result as the so-called Pierce Lemma established in [7] (see also [5]) for Voronoi quantization to prove the original Zador Theorem.

In the proof, we will make use of the Γ and B functions defined by $\Gamma(a) = \int_0^{+\infty} u^{a-1}e^{-u}du$, $a > 0$, and $B(a, b) = \int_0^1 u^{a-1}(1-u)^{b-1}du$, $a, b > 0$, respectively, and satisfying $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

Proof. Step 1. We first assume that X is $[1, +\infty)$-valued and $n \geq 2$. Let $(Y_n)_{n\geq 1}$ be a sequence of i.i.d. Pareto(δ)-distributed random variables (with probability density $f_Y(y) = \delta y^{-\delta-1}\mathbf{1}_{\{y\geq 1\}}$) defined on a probability space $(\Omega', \mathcal{A}', \mathbb{P}')$.

Let $\delta = \delta(\eta, p) \in (0, \frac{\eta}{\lceil p \rceil})$ be chosen so that $\ell = \ell(p, \eta) = \frac{p}{\delta}$ is an integer and $\ell \geq 2$. For every $n \geq \ell(p, \eta)$, set $\widetilde{n} = n - \ell + 2 \in \mathbb{N}$, $\widetilde{n} \leq n$. It follows from the monotony property (18) that

$$\begin{aligned}
\inf_{(1,x_2,\ldots,x_n)\in\mathcal{I}_n} \|\bar{F}_{p,n}(X, 1, x_2, \ldots, x_n)\|_{L^p} &\leq \inf_{(1,x_2,\ldots,x_{\widetilde{n}})\in\mathcal{I}_{\widetilde{n}}} \|\bar{F}_{p,\widetilde{n}}(X, 1, x_2, \ldots, x_{\widetilde{n}})\|_{L^p}\\
&\leq \|\bar{F}_{p,\widetilde{n}}(X, Y_0^{(n)}, Y_1^{(n)}, \ldots, Y_{\widetilde{n}-1}^{(n)})\|_{L^p(\Omega\times\Omega',\mathbb{P}\otimes\mathbb{P}')}
\end{aligned}$$

where, for every $n \geq 1$, $Y^{(n)} = (Y_1^{(n)}, \ldots, Y_n^{(n)})$ denotes the standard order statistics of the first n terms of the sequence $(Y_k)_{k\geq 1}$ and $Y_0^{(n)} = 1$. On the other hand, we recall (see *e.g.* [3]) that the joint distribution of $(Y_i^{(n)}, Y_{i+1}^{(n)})$, $1 \leq i \leq n-1$, is given by

$$\mathbb{P}'_{(Y_i^{(n)}, Y_{i+1}^{(n)})}(du, dv) = \delta^2 \frac{n!}{(i-1)!(n-i-1)!}(1 - u^{-\delta})^{i-1}v^{-\delta(n-i-1)}(uv)^{-\delta-1}\, du\, dv.$$

Step 2. Assume that $n \geq 3$. Since X and $(Y_1, \ldots, Y_0)$ are independent and $X \geq 1$

$$\|\bar{F}_{p,\widetilde{n}}(X, Y_0^{(n)}, Y_1^{(n)}, \ldots, Y_{\widetilde{n}-1}^{(n)})\|^p_{L^p(\Omega\times\Omega',\mathbb{P}\otimes\mathbb{P}')} = \int_{[1,+\infty)} \|\bar{F}_{p,\widetilde{n}}(\xi, Y_0^{(n)}, Y_1^{(n)}, \ldots, Y_{\widetilde{n}-1}^{(n)})\|^p_{L^p(\Omega',\mathbb{P}')}\mathbb{P}_X(d\xi).$$

Relying on the expression (16) of the functional $\bar{F}_n^p$, we set for every $i = 0, \ldots, n-\ell$ and $\xi \geq 1$

$$(a)_i := \mathbb{E}\left(\frac{(Y_{i+1}^{(n)} - \xi)^p(\xi - Y_i^{(n)})}{Y_{i+1}^{(n)} - Y_i^{(n)}}\mathbf{1}_{\{Y_i^{(n)} < \xi \leq Y_{i+1}^{(n)}\}}\right),\ (b)_i := \mathbb{E}\left(\frac{(Y_{i+1}^{(n)} - \xi)(\xi - Y_i^{(n)})^p}{Y_{i+1}^{(n)} - Y_i^{(n)}}\mathbf{1}_{\{Y_i^{(n)} < \xi \leq Y_{i+1}^{(n)}\}}\right)$$

and $(c)_{\widetilde{n}-1} := \mathbb{E}\Big(\big(\xi - Y_{n-\ell+1}^{(n)}\big)^p\mathbf{1}_{\{\xi\geq Y_{n-\ell+1}^{(n)}\}}\Big)$.

We will first inspect the sum $\sum_{i=0}^{n-\ell}(\square)_i$, $\square = a, b$ successively.

Let $i \in \{1, \ldots, \tilde{n} - 1\}$. It follows from the above expression of the distribution of $(Y_i^{(n)}, Y_{i+1}^{(n)})$ that

$$(a)_i = \delta^2 \iint_{1\leq u\leq\xi\leq v} \frac{(v-\xi)^p(\xi-u)}{v-u}(1-u^{-\delta})^{i-1}v^{-\delta(n-i-1)}(uv)^{-\delta-1}\, du\, dv \frac{n!}{(i-1)!(n-i-1)!}.$$

The change of variable $v = \xi(w+1)$ yields

$$(a)_i = n(n-1)\binom{n-2}{i-1}\delta^2 \int_1^\xi du\, (\xi-u)(1-u^{-\delta})^{i-1}u^{-\delta-1}\xi^{p-\delta(n-i)}\int_0^{+\infty} dw\, \frac{w^p}{\xi(w+1)-u}(w+1)^{-\delta(n-i)-1}.$$

Noting that $\frac{\xi-u}{\xi(w+1)-u} \leq \frac{1}{w+1}$ then leads to

$$(a)_i \leq n(n-1)\binom{n-2}{i-1}\delta^2 n(n-1)\xi^{p-\delta(n-i)}\int_1^{\xi}(1-u^{-\delta})^{i-1}u^{-\delta-1}du \times \int_0^{+\infty} w^p(1+w)^{-\delta(n-i)-2}dw.$$

The change of variable $w=\frac{1}{y}-1$ shows that $\displaystyle\int_0^{+\infty} w^p(1+w)^{-\delta(n-i)-2}dw = B(\delta(n-i)-p+1, p+1)$ whereas $\displaystyle\int_1^{\xi}(1-u^{-\delta})^{i-1}u^{-\delta-1}du = \frac{(1-\xi^{-\delta})^i}{\delta i}$ so that

$$(a)_i \leq \delta n\binom{n-1}{i}(1-\xi^{-\delta})^i \xi^{p-\delta(n-i)}\frac{\Gamma(p+1)\Gamma(\delta(n-i)-p+1)}{\Gamma(\delta(n-i)+2)}$$

where we used the standard identity $\binom{n-1}{i} = \frac{n-1}{i}\binom{n-2}{i-1}$.

When $i=0$, noting that the density of $Y_1^{(n)} = \min_{1\leq i\leq n} Y_i$ is $\delta n y^{-\delta n-1}\mathbf{1}_{\{y\geq 1\}}$, we get

$$\begin{aligned}
(a)_0 &= \mathbb{E}\left(\frac{(Y_1^{(n)}-\xi)^p(\xi-1)}{Y_1^{(n)}-1}\mathbf{1}_{\{1\leq\xi\leq Y_1^{(n)}\}}\right)\\
&= \delta n\int_{\xi}^{+\infty}(\xi-1)\frac{(v-\xi)^p}{v-1}v^{-\delta n-1}dv\\
&= \delta n\xi^{p-\delta n}\int_0^{+\infty}\frac{(\xi-1)}{\xi(w+1)-1}w^p(w+1)^{-\delta n-1}dw \quad \text{where we set } v=\xi(w+1)\\
&\leq \delta n\xi^{p-\delta n}B(\delta n-p+1,p+1)
\end{aligned}$$

where we used in the last line that $\frac{\xi-1}{\xi(w+1)-1} \leq \frac{1}{w+1}$. As a consequence

$$\begin{aligned}
\sum_{i=0}^{n-\ell}(a)_i &\leq \delta\, n\,\Gamma(p+1)\sum_{i=0}^{n-\ell}\binom{n-1}{i}\xi^{p-\delta(n-i)}(1-\xi^{-\delta})^i\frac{\Gamma(\delta(n-i)-p+1}{\Gamma(\delta(n-i)+2}\\
&\leq \delta\, n\,\Gamma(p+1)\xi^p(1-\xi^{-\delta})^n\sum_{j=\ell}^{n}\binom{n-1}{j-1}(\xi^{\delta}-1)^{-j}\frac{\Gamma(\delta j-p+1)}{\Gamma(\delta j+2)}.
\end{aligned}$$

Now using that for every $a>0$, $\frac{\Gamma(x+a)}{\Gamma(x)} \sim x^a$ as $x\to\infty$, we derive the existence of a real constants $\tilde{\kappa}^{(0)}_{p,\delta}$, $\kappa^{(0)}_{p,\delta}>0$ such that

$$\forall\, j\geq 0,\qquad \frac{\Gamma(\delta j-p+1)}{\Gamma(\delta j+2)} \leq \tilde{\kappa}^{(0)}_{p,\delta}\, j^{-(p+1)} \leq \kappa^{(0)}_{p,\delta}\,\frac{j^{\lceil p\rceil-p}}{j(j+1)\cdots(j+\lceil p\rceil)}.$$

In turn, using that

$$\binom{n+\lceil p\rceil}{j+\lceil p\rceil} = \frac{(n+\lceil p\rceil)\cdots n}{(j+\lceil p\rceil)\cdots j}\binom{n-1}{j-1},$$

we finally obtain

$$\begin{aligned}
\sum_{i=0}^{n-\ell}(a)_i &\leq \kappa^{(0)}_{p,\delta}\, n\Gamma(p+1)\xi^p\delta(1-\xi^{-\delta})^n\frac{1}{(n+\lceil p\rceil)\cdots(n+1)n}\sum_{j=\ell}^{n}\binom{n+\lceil p\rceil}{j+\lceil p\rceil}(\xi^{\delta}-1)^{-j}j^{\lceil p\rceil-p}\\
&\leq \kappa^{(0)}_{p,\delta}\,\Gamma(p+1)\xi^p\delta(1-\xi^{-\delta})^n\frac{n^{\lceil p\rceil-p}}{(n+\lceil p\rceil)\cdots(n+1)}(\xi^{\delta}-1)^{\lceil p\rceil}\Big(1+(\xi^{\delta}-1)^{-1}\Big)^{n+\lceil p\rceil}.
\end{aligned}$$

Now

$$(1-\xi^{-\delta})^n\xi^p(\xi^\delta-1)^{\lceil p\rceil}\Big(1+(\xi^\delta-1)^{-1}\Big)^{n+\lceil p\rceil}=\xi^{p+\delta\lceil p\rceil}$$

so that, using that $\xi\geq 1$ and $\delta<\frac{\eta}{\lceil p\rceil}$, we get $\xi^{p+\delta\lceil p\rceil}\leq \xi^{p+\eta}$ which in turn implies

$$\sum_{i=0}^{n-\ell}(a)_i\leq \kappa^{(0)}_{p,\delta}\,\delta\,\Gamma(p+1)\xi^{p+\eta}\frac{1}{n^p}.$$

Let us pass now to the second sum involving $(b)_i$. First note that, on the event $\Big\{Y_i^{(n)}\leq\xi\leq\dfrac{Y_i^{(n)}+Y_{i+1}^{(n)}}{2}\Big\}$ (which is clearly included in $\{Y_i^{(n)}\leq\xi\leq Y_{i+1}^{(n)}\}$), one has $(\xi-Y_i^{(n)})^p(Y_{i+1}^{(n)}-\xi)\leq(\xi-Y_i^{(n)})(Y_{i+1}^{(n)}-\xi)^p$ so that, owing to what precedes, we can focus on $\displaystyle\sum_{i=0}^{n-\ell}(\widetilde{b})_i$ where

$$(\widetilde{b})_i:=\mathbb{E}\left((\xi-Y_i^{(n)})^p\mathbf{1}_{\big\{\frac{Y_i^{(n)}+Y_{i+1}^{(n)}}{2}\leq\xi\leq Y_{i+1}^{(n)}\big\}}\right)\geq\mathbb{E}\left(\frac{(Y_{i+1}^{(n)}-\xi)(\xi-Y_i^{(n)})^p}{Y_{i+1}^{(n)}-Y_i^{(n)}}\mathbf{1}_{\big\{\frac{Y_i^{(n)}+Y_{i+1}^{(n)}}{2}\leq\xi\leq Y_{i+1}^{(n)}\big\}}\right).$$

This time we will analyze successively the sum over $i=1,\ldots,n-\ell$ and the case $i=0$.

$$\begin{aligned}
\sum_{i=1}^{n-\ell}(\widetilde{b})_i &= \delta^2n(n-1)\iint_{\{1\leq u\leq\xi\leq v\leq 2\xi-u\}}du\,dv\,(uv)^{-\delta-1}(\xi-u)^p\sum_{i=1}^{n-\ell}\binom{n-2}{i-1}v^{-\delta(n-2-(i-1))}(1-u^{-\delta})^{i-1}\\
&\leq \delta^2n(n-1)\iint_{\{1\leq u\leq\xi\leq v\leq 2\xi-u\}}du\,dv(uv)^{-\delta-1}(\xi-u)^p(1-u^{-\delta}+v^{-\delta})^{n-2}\\
&\leq \delta^2n(n-1)\int_1^\xi du u^{-\delta-1}(\xi-u)^p\int_\xi^{2\xi-u}dv\,v^{-\delta-1}e^{-(n-2)(u^{-\delta}-v^{-\delta})}\\
&= \delta^2n(n-1)\int_1^\xi du u^{-\delta-1}(\xi-u)^pe^{-(n-2)u^{-\delta}}\int_\xi^{2\xi-u}dv\,v^{-\delta-1}e^{(n-2)v^{-\delta}}
\end{aligned}$$

where we used in the in the second line that $n-\ell-1\leq n-2$ since $\ell\geq 1$. Setting $v=y^{-\frac{1}{\delta}}$ yields

$$\begin{aligned}
\int_\xi^{2\xi-u}v^{-\delta-1}e^{(n-2)v^{-\delta}}\,dv &= \frac{1}{\delta}\int_{(2\xi-u)^{-\delta}}^{\xi^{-\delta}}e^{(n-2)y}dy\\
&\leq \frac{1}{\delta}\big(\xi^{-\delta}-(2\xi-u)^{-\delta}\big)e^{(n-2)\xi^{-\delta}}\\
&\leq (\xi-u)\xi^{-\delta-1}e^{(n-2)\xi^{-\delta}}
\end{aligned}$$

where we used in the last line the fundamental formula of Calculus. Consequently,

$$\begin{aligned}
\sum_{i=1}^{n-\ell}(\widetilde{b})_i &\leq n(n-1)\delta^2\xi^{-\delta-1}\int_1^\xi u^{-\delta-1}(\xi-u)^{p+1}e^{-(n-2)(u^{-\delta}-\xi^{-\delta})}du\\
&= n(n-1)\xi^{-\delta-1}\delta\int_0^{(n-2)(1-\xi^{-\delta})}\Big(\xi-\big(\frac{x}{n-2}+\xi^{-\delta}\big)^{-\frac{1}{\delta}}\Big)^{p+1}e^{-x}\frac{dx}{n-2}
\end{aligned}$$

where we put $u=\big(\frac{x}{n-2}+\xi^{-\delta}\big)^{-\frac{1}{\delta}}$. Now, applying again fundamental formula of Calculus to the function $z^{-\frac{1}{\delta}}$ yields,

$$\xi-\Big(\frac{x}{n-2}+\xi^{-\delta}\Big)^{-\frac{1}{\delta}}=(\xi^{-\delta})^{-\frac{1}{\delta}}-\Big(\frac{x}{n-2}+\xi^{-\delta}\Big)^{-\frac{1}{\delta}}\leq\frac{x}{\delta(n-2)}\xi^{\delta+1}$$

$$\text{so that}\quad \sum_{i=1}^{n-\ell}(\widetilde{b})_i \;\le\; \frac{n(n-1)}{(n-2)^{p+2}}\delta^{-p}\xi^{(p+1)(\delta+1)-(\delta+1)}\int_0^{(n-2)(1-\xi^{-\delta})} x^{p+1}e^{-x}dx$$
$$\le\; \kappa^{(1)}_{p,\delta}\Gamma(p+2)n^{-p}\xi^{p(\delta+1)}$$

for some constant $\kappa^{(1)}_{p,\delta}>0$.

When $i=0$, keeping in mind that $Y^{(n)}_1=\min_{1\le i\le n}Y_i$,

$$\begin{aligned}(\widetilde{b})_0 &\le (\xi-1)^p\mathbb{P}(\xi\le Y^{(n)}_1\le 2\xi-1)=(\xi-1)^p\big(\xi^{-n\delta}-(2\xi-1)^{-n\delta}\big)\\ &\le n\delta(\xi-1)^{p+1}\xi^{-n\delta-1}=n\delta\xi^{p(1+\delta)}g(1/\xi)\end{aligned}$$

where $g(u)=(1-u)^{p+1}u^{(n+p)\delta}$, $u\in(0,1)$. One checks that g attains its maximum over $(0,1]$ at $u^*=\frac{(n+p)\delta}{(n+p)\delta+p+1}$ so that

$$\sup_{u\in(0,1]}g(u)=g(u^*)=\left(\frac{p+1}{(n+p)\delta+p+1}\right)^{p+1}(u^*)^{(n+p)\delta}\le\left(\frac{1}{1+\frac{n+p}{p+1}\delta}\right)^{p+1}.$$

Finally, there exists a real constant $\kappa^{(2)}_{p,\delta}>0$ such that

$$(\widetilde{b})_0\le\xi^{p(\delta+1)}\frac{\delta n}{(1+\frac{n+p}{p+1}\delta)^{p+1}}\le\kappa^{(2)}_{p,\delta}\xi^{p(\delta+1)}n^{-p}.$$

As concerns the $(c)_{n-\ell+1}$ term, we proceed as follows.

$$\begin{aligned}\mathbb{E}\Big(\big(\xi-Y^{(n)}_{n-\ell+1}\big)^p\mathbf{1}_{\{\xi\ge Y^{(n)}_{n-\ell+1}\}}\Big) &\le \xi^p\mathbb{P}(\xi\ge Y^{(n)}_{n-\ell+1})\\ &\le \xi^{p(1+\delta)}\mathbb{E}(Y^{(n)}_{n-\ell+1})^{-p\delta}.\end{aligned}$$

Note that

$$\begin{aligned}\mathbb{E}(Y^{(n)}_{n-\ell+1})^{-p\delta} &= \frac{\Gamma(n+1)}{\Gamma(n-\ell+1)\Gamma(\ell)}\int_0^1(1-v)^{n-\ell}v^{\ell+p-1}dv=\frac{\Gamma(n+1)}{\Gamma(\ell)}\frac{\Gamma(\ell+p)}{\Gamma(n+p+1)}\\ &\sim \frac{\Gamma(\ell+p)}{\Gamma(\ell)}n^{-p}=O(n^{-p}).\end{aligned}$$

Finally, for every $\xi\ge1$,

$$(c)_{n-\ell+1}\le\kappa^{(3)}_{p,\delta}\xi^{p(1+\delta)}n^{-p}.$$

Consequently, there exists a real constant $\kappa_{p,\eta}=\max_{j=0,\dots,3}\kappa^{(j)}_{p,\delta}>0$ such that for every $n\ge n_{p,\eta}=\ell(\eta,p)\vee3$,

$$\forall\,\xi\ge1,\qquad n^p\,\mathbb{E}\,\bar F^p_n(\xi,Y^{(n)}_0,\dots,Y^{(n)}_{\tilde n+1})\le\kappa_{p,\eta}\,\xi^{p+\eta}$$

since $p\,\delta\le\eta$. Hence for every r.v. X, we derive by integrating in $\xi\in[1,+\infty)$ with respect to $\mathbb{P}_X(d\xi)$:

$$n^p\inf_{(1,x_2,\dots,x_n)\in\mathcal{I}_n}\mathbb{E}\,\bar F^p_n(X,1,x_2,\dots,x_n)\le n^p\mathbb{E}\,\bar F^p_n(X,Y^{(n)}_0,\dots,Y^{(n)}_{\tilde n+1})\le\kappa_{p,\eta}\,\mathbb{E}\,X^{p+\eta}.$$

Step 3. If X is a non-negative random variable, applying the second step to $X+1$ and using the scaling property (i) satisfied by $F_{p,n}$ yields for $n\ge n_{p,\eta}$ (as defined in Step 2),

$$\begin{aligned}\inf_{(0,x_2,\dots,x_n)\in\mathcal{I}_n}\|\bar F_{p,n}(X,0,x_2,\dots,x_n)\|_{L^p} &= \inf_{(1,x_2,\dots,x_n)\in\mathcal{I}_n}\|\bar F_{p,n}(X+1,1,\dots,x_n)\|_{L^p}\\ &\le \kappa^{1/p}_{p,\eta}\frac{\|1+X\|^{1+\frac{\eta}{p}}_{L^{p+\eta}}}{n}\\ &\le C^{(0)}_{p,\eta}\frac{(1+\|X\|^{1+\frac{\eta}{p}}_{L^{p+\eta}})}{n}\quad\text{with } C^{(0)}_{p,\eta}=(2^{1+\eta}\kappa_{p,\eta})^{\frac1p}.\end{aligned}$$

We may assume that $\|X\|_{L^{p+\eta}} \in (0,\infty)$. Then, applying the above bound to the non-negative random variable $\widetilde{X} = \frac{X}{\|X\|_{L^{p+\eta}}}$ taking again advantage of the scaling property (i), we obtain

$$\begin{aligned}\inf_{(0,x_2,\dots,x_n)\in\mathcal{I}_n} \|\bar{F}_{p,n}(X,0,x_2,\dots,x_n)\|_{L^p} &= \|X\|_{L^{p+\eta}} \inf_{(0,x_2,\dots,x_n)\in\mathcal{I}_n} \|\bar{F}_{p,n}(\widetilde{X},0,x_2,\dots,x_n)\|_{L^p}\\ &\le \|X\|_{L^{p+\eta}} C^{(0)}_{p,\eta}\frac{1+1}{n} = 2C^{(0)}_{p,\eta}\|X\|_{L^{p+\eta}}\frac{1}{n}.\end{aligned}$$

STEP 4. Let X be a real-valued random variable and let for every integer $n \ge 1$, $x_1,\dots,x_n \in (-\infty,0)$, $x_{n+1} = 0$ and $x_{n+2},\dots,x_{2n+1} \in (0,+\infty)$. It follows from the additivity property that that

$$\begin{aligned}\bar{F}^p_{2n+1}(X,x_1,\dots,x_{2n+1}) &= \bar{F}^p_{n+1}(X_+,x_{n+1},\dots,x_{2n+1})\mathbf{1}_{\{X\ge 0\}}\\ &\quad+\bar{F}^p_{n+1}(-X_-,x_1,\dots,x_{n+1})\mathbf{1}_{\{X<0\}}\\ &= \bar{F}^p_{n+1}(X_+,x_1,\dots,x_{n+1})\mathbf{1}_{\{X\ge 0\}} + \bar{F}^p_{n+1}(X_-,-x_{n+1},\dots,-x_1)\mathbf{1}_{\{X<0\}}.\end{aligned}$$

Consequently, using that $X_+ \times X_- \equiv 0$ and that $x_{n+1} = 0$, we get

$$\begin{aligned}\inf_{\substack{(x_1,\dots,x_{2n+1})\in\mathcal{I}_{2n+1}\\ x_{n+1}=0}} \|\bar{F}_{p,2n+1}(X,x_1,\dots,x_{2n+1})\|^p_{L^p} \le &\inf_{(0,y_2,\dots,y_{n+1})\in\mathcal{I}_{n+1}} \|\bar{F}_{p,n+1}(X_+,0,y_2,\dots,y_{n+1})\|^p_{L^p}\\ &+\inf_{(0,y_2,\dots,y_{n+1})\in\mathcal{I}_{n+1}} \|\bar{F}_{p,n}(X_-,0,y_2,\dots,y_{n+1})\|^p_{L^p}.\end{aligned}$$

Hence, it follows from Step 2 that, for every $n \ge n_{p,\eta}-1$,

$$\inf_{(x_1,\dots,x_{2n+1})\in\mathcal{I}_{2n+1}} \|\bar{F}_{p,2n+1}(X,x_1,\dots,x_{2n+1})\|^p_{L^p} \le \Big(\|X_-\|^p_{L^{p+\eta}} + \|X_+\|^p_{L^{p+\eta}}\Big)\Big(\frac{2C^{(0)}_{p,\eta}}{n+1}\Big)^p.$$

Now using that $(a+b) \le 2^{1-\frac{1}{q}}(a^q+b^q)^{\frac{1}{q}}$, $a,b\ge 0$, with $q = 1+\frac{\eta}{p} \ge 1$, we derive that

$$\|X_-\|^p_{L^{p+\eta}} + \|X_+\|^p_{L^{p+\eta}} \le 2^{\frac{\eta}{p+\eta}}\Big(\|X_-\|^{p+\eta}_{L^{p+\eta}} + \|X_+\|^{p+\eta}_{L^{p+\eta}}\Big)^{\frac{p}{p+\eta}} = 2^{\frac{\eta}{p+\eta}}\|X\|^p_{L^{p+\eta}}$$

since $X_- \times X_+ \equiv 0$. Now, the monotonicity property (17) implies that, for every $n \ge 2\,n_{p,\eta}$,

$$\bar{d}_{n,p}(X) = \inf_{(x_1,\dots,x_n)\in\mathcal{I}_n} \|\bar{F}_{p,n}(X,x_1,\dots,x_n)\|_{L^p} \le 2^{\frac{\eta}{p(p+\eta)}}2C^{(0)}_{p,\eta}\frac{\|X\|_{L^{p+\eta}}}{n}.$$

Still calling upon (17), we note that, for every $n\in\{1,\dots,2n_{p,\eta}\}$, $\bar{d}_{n,p}(X) \le \bar{d}_{1,p}(X) = \inf_{x\in\mathbb{R}}\|X-x_1\|_{L^p} \le \|X\|_{L^p}$ so that

$$\bar{d}_{n,p}(X) \le 2n_{p,\eta}\frac{\|X\|_{L^{p+\eta}}}{n}$$

which completes the proof by setting $C_{p,\eta} = \max\big(2n_{p,\eta}, 2^{1+\frac{\eta}{p(p+\eta)}}C^{(0)}_{p,\eta}\big)$. □

4.1 A d-dimensional non-asymptotic upper-bound for the dual quantization error

Now, combining Theorem 5 and Proposition 4(b), we are in position to show Proposition 2 (the d-dimensional version of the extended Pierce Lemma) which provides a non-asymptotic upper-bound at the exact rate for dual quantization error moduli.

Proof of Proposition 2. (a) First note that $\bar{d}_{n,p}(X) = \bar{d}_{n,p}(X-a)$, $a\in\mathbb{R}^d$ (invariance by translation) so we may assume that X is $L^{p+\eta}$-centered *i.e.* $\sigma_{p+\eta,\|.\|}(X) = \|X\|_{L^{p+\eta}}$. When $d = 1$, Theorem 5 solves the problem.

Let $d \geq 2$. Let $X = (X^1, \ldots, X^d)$ (X^i components of X). It follows form Proposition 4 that, if $\Gamma = \prod_{1 \leq i \leq d} \Gamma_i$, with $\Gamma_i \subset \mathbb{R}$, $|\Gamma_i| = n_i$ with $n_1 \cdots n_d \leq n$. Then for every $\xi = (\xi^1, \ldots, \xi^d) \in \mathbb{R}^d$

$$\bar{F}^p_{\|.\|}(\xi;\Gamma) \leq C_{p,\|.\|} \bar{F}^p_{\ell^p}(\xi;\Gamma) = \sum_{j=1}^{d} \bar{F}^p(\xi^j, \Gamma_j)$$

where $C_{p,\|.\|} = \sup_{|\xi|_{\ell^p}=1} \|\xi\|^p$. Integrating with respect to the distribution of X yields $\bar{d}^p(X,\Gamma) \leq C_{p,\|.\|} \sum_{j=1}^{d} \bar{d}^p(X^j, \Gamma_j)$ which in turn easily implies

$$\bar{d}^p_n(X) \leq C_{p,\|.\|} \sum_{j=1}^{d} \bar{d}^p_{n_j}(X^j).$$

Now set $n_j = \lfloor n^{\frac{1}{d}} \rfloor$, $j = 1, \ldots, d$. It follows from Theorem 5 that

$$\begin{aligned}
\bar{d}^p_n(X) &\leq C^p_{p,\|.\|} C_{p,\eta} \sum_{j=1}^{d} \|X^j\|^p_{L^{p+\eta}} \lfloor n^{\frac{1}{d}} \rfloor^{-p} \\
&\leq C_{p,\|.\|} C_{p,\eta} \sup_{k \geq 2} \left(\frac{k^{\frac{1}{d}}}{k^{\frac{1}{d}} - 1} \right)^p n^{-\frac{p}{d}} \sum_{j=1}^{d} \|X^j\|^p_{L^{p+\eta}} \\
&\leq C_{p,\|.\|} C_{p,\eta} 2^p n^{-\frac{p}{d}} d^{\frac{\eta}{p+\eta}} \mathbb{E}|X|^{p+\eta}_{\ell^{p+\eta}} \\
&\leq d^{\frac{\eta}{p+\eta}} C_{p,\|.\|} C_{p,\eta}\, 2^p \widetilde{C}_{\|.\|,p+\eta} \|X\|^{p+\eta}_{L^{p+\eta}} n^{-\frac{p}{d}}
\end{aligned}$$

where $\widetilde{C}_{\|.\|,r} = \sup_{\|x\|=1} |x|^r_{\ell^r}$, $r > 0$.

(b) Let C be the smallest hypercube withe edges parallel to the coordinate axis containing $\mathrm{conv}(\mathrm{Supp}(\mathbb{P}_X))$. Up to a translatation, which leaves $d_{n,p}(X)$ invariant, we may assume that $C = [0, L]^d$ where $0 \leq L \leq \mathrm{diam}_{\|.\|}(\mathrm{Supp}(\mathbb{P}_X))$. The conclusion follows by integrating Inequality (15) with respect to $\mathbb{P}_X(d\xi)$ with $m = \lfloor n^{\frac{1}{d}} \rfloor$ and following the lines of the proof of claim (a). $\square$

5 Proof of the sharp rate theorem

On the way to proving the sharp rate theorem, we have to establish few additional propositions.

Proposition 5 (Sub-linearity)**.** *Let $\mathbf{P} = \sum_{i=1}^m s_i \mathbf{P}_i$ where $s_1, \ldots, s_m \in [0,1]$, $\sum_{i=1}^m s_i = 1$ and let $n_1, \ldots, n_m \in \mathbb{N}$ such that $\sum_{i=1}^m n_i \leq n$. Then*

$$d^p_n(\mathbf{P}) \leq \sum_{i=1}^{m} s_i\, d^p_{n_i}(\mathbf{P}_i).$$

Proof. For $\varepsilon > 0$ and every $i = 1, \ldots, m$, choose $\Gamma_i \subset \mathbb{R}^d$, $|\Gamma_i| \leq n_i$ such that

$$d^p(\mathbf{P}_i; \Gamma_i) \leq (1+\varepsilon)\, d^p_{n_i}(\mathbf{P}_i).$$

Set $\Gamma = \bigcup_{i=1}^m \Gamma_i$; from Proposition 3 we get

$$\begin{aligned}
d^p_n(\mathbf{P}) \leq d^p_n(\mathbf{P};\Gamma) &= \sum_{i=1}^{m} s_i \int F^p(\xi;\Gamma)\, \mathbf{P}_i(d\xi) \\
&\leq \sum_{i=1}^{m} s_i \int F^p(\xi;\Gamma_i)\, \mathbf{P}_i(d\xi) \leq (1+\varepsilon) \sum_{i=1}^{m} s_i\, d^p_{n_i}(\mathbf{P}_i).
\end{aligned}$$

Letting $\varepsilon \to 0$ completes the proof. $\square$

Remark. Proposition 5 does not hold for $\bar{d}_n^p$ since $\bar{F}^p$ is not decreasing for the inclusion order on grids. This induces substantial difficulties in the proof of the sharp rate compared to the regular quantization setting.

Proposition 6 (Scaling property)**.** *Let $C = a + \rho[0,1]^d$ ($a \in \mathbb{R}^d$, $\rho > 0$) be a d-dimensional hypercube, with edges parallel to the coordinate axis and edge-length $\rho > 0$. Then*

$$d_{n,p}(\mathcal{U}(C)) = \rho \cdot d_{n,p}(\mathcal{U}\big([0,1]^d\big)).$$

Proof. Keeping in mind that $\lambda_d([0,\rho]^d) = \rho^d$, it holds that

$$\begin{aligned}
d^p(\mathcal{U}(C); \{a+\rho x_1, \ldots, a+\rho x_n\}) &= \int_{[0,\rho]^d} \min_{\substack{\lambda\in\mathbb{R}^n \\ \text{s.t. } \left[\begin{smallmatrix} \rho x_1 & \cdots & \rho x_n \\ 1 & \cdots & 1\end{smallmatrix}\right]\lambda = \left[\begin{smallmatrix}\xi \\ 1\end{smallmatrix}\right],\, \lambda\geq 0}} \sum_{i=1}^n \lambda_i \, \|\xi - \rho x_i\|^p \; \frac{\lambda_d(d\xi)}{\lambda_d\big([0,\rho]^d\big)} \\
&= \int_{[0,1]^d} \min_{\substack{\lambda\in\mathbb{R}^n \\ \text{s.t. } \left[\begin{smallmatrix} \rho x_1 & \cdots & \rho x_n \\ 1 & \cdots & 1\end{smallmatrix}\right]\lambda = \left[\begin{smallmatrix}\rho u \\ 1\end{smallmatrix}\right],\, \lambda\geq 0}} \sum_{i=1}^n \lambda_i \, \|\rho u - \rho x_i\|^p \; \lambda_d(du) \\
&= \rho^p \int_{[0,1]^d} \min_{\substack{\lambda\in\mathbb{R}^n \\ \text{s.t. } \left[\begin{smallmatrix} x_1 & \cdots & x_n \\ 1 & \cdots & 1\end{smallmatrix}\right]\lambda = \left[\begin{smallmatrix}u \\ 1\end{smallmatrix}\right],\, \lambda\geq 0}} \sum_{i=1}^n \lambda_i \, \|u - x_i\|^p \; \lambda_d(du) \\
&= \rho^p \cdot d^p(\mathcal{U}\big([0,1]^d\big); \{x_1, \ldots, x_n\}).
\end{aligned}$$

$\square$

The following lemma shows that also for $\bar{d}_{n,p}$ the convex hull spanned by a sequence of "semi-optimal" quantizers asymptotically covers the interior of $\text{supp}(\mathbb{P}_X)$. This fact is trivial for $d_{n,p}$ if X has a compact support.

Lemma 1. *Let $K = \text{conv}\{a_1, \ldots, a_k\} \subset \overbrace{\text{supp}(\mathbf{P})}^{\circ}$ be a set with $\mathring{K} \neq \emptyset$ and let Γ_n be a sequence of quantizers such that $\bar{d}_{n,p}(\mathbf{P}, \Gamma_n) \to 0$ as $n \to +\infty$. Then there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$*

$$K \subset \text{conv}(\Gamma_n).$$

Proof. Set $a_0 = \frac{1}{k}\sum_{i=1}^k a_i$ and define for every $\rho > 0$

$$\tilde{K}(\rho) = \text{conv}\{\tilde{a}_1(\rho), \ldots, \tilde{a}_k(\rho)\} \quad \text{with} \quad \tilde{a}_i(\rho) = a_0 + (1+\rho)(a_i - a_0).$$

Since $K \subset \overbrace{\text{supp}(\mathbf{P})}^{\circ}$ there exists $\rho_0 > 0$ such that $\tilde{K} = \tilde{K}(\rho_0) \subset \text{supp}(\mathbf{P})$. From now on, we denote $\tilde{a}_i(\rho_0)$ by $\tilde{a}_i$. Since moreover $\tilde{a}_i \in \text{supp}(\mathbf{P})$, there exists a sequence $(a_i^n)_{n\geq 1}$ having values in $\text{conv}(\Gamma_n)$ and converging to $\tilde{a}_i$. Otherwise there would exist $\varepsilon_0 > 0$ and a subsequence (n') such that $B(\tilde{a}_i, \varepsilon_0) \subset (\text{conv}(\Gamma_{n'}))^c$. Then

$$\bar{d}_{n'}^p(X, \Gamma_{n'}) \geq \mathbb{E}\,\text{dist}(X, \Gamma_{n'})^p \mathbf{1}_{\{X\in B(\tilde{a}_i, \varepsilon_0/2)\}} \left(\frac{\varepsilon_0}{2}\right)^p \mathbf{P}(B(\tilde{a}_i, \varepsilon_0/2)) > 0$$

since $\tilde{a}_i \in \text{supp}(\mathbf{P})$. This contradicts the assumption on the sequence $(\Gamma_n)_{n\geq 1}$.

Since K has a nonempty interior, it follows that $\text{aff.dim}\{a_1, \ldots, a_k\} = \text{aff.dim}\{\tilde{a}_1, \ldots, \tilde{a}_k\} = d$. Consequently, we may choose a subset $I^* \subset \{1, \ldots, k\}$, $|I^*| = d+1$, so that $\{\tilde{a}_j : j \in I^*\}$ is an affinely independent system in $\mathbb{R}^d$ and furthermore there exists $n_0 \in \mathbb{N}$ such that the same holds for $\{a_j^n : j \in I^*\}$, $n \geq n_0$. Hence, we may write for $n \geq n_0$

$$\tilde{a}_i = \sum_{j\in I^*} \mu_j^{n,i} a_j^n, \quad \sum_{j\in I^*} \mu_j^{n,i} = 1, \quad i = 1, \ldots, k. \tag{19}$$

This linear system has the unique asymptotic solution $\mu_j^{\infty,i} = \delta_{ij}$ (Kronecker symbol), which implies $\mu_j^{n,i} \to \delta_{ij}$ when $n \to +\infty$.

Now let $\xi \in K \subset \tilde{K}$ and write

$$\xi = \sum_{i=1}^{k} \lambda_i a_i \quad \text{for some} \quad \lambda_i \geq 0, \sum_{i=1}^{k} \lambda_i = 1.$$

One easily checks that it also holds

$$\xi = \sum_{i=1}^{k} \tilde{\lambda}_i \tilde{a}_i \;\text{ with }\; \tilde{\lambda}_i = \frac{\rho_0}{k(1+\rho_0)} + \frac{\lambda_i}{1+\rho_0} \geq \frac{\rho_0}{k(1+\rho_0)} > 0 \quad \text{and} \quad \sum_{i=1}^{k} \tilde{\lambda}_i = 1.$$

Furthermore, we may choose $n_1 \geq n_0$ such that, for every $n \geq n_1$,

$$\mu_i^{n,i} > \frac{1}{2} \quad \text{and} \quad \forall j \neq i,\ |\mu_j^{n,i}| \leq \frac{\rho_0}{4k(1+\rho_0)}.$$

Using (19), this leads to

$$\xi = \sum_{j \in I^*} \Big(\sum_{i=1}^{k} \tilde{\lambda}_i \mu_j^{n,i} \Big) a_j^n$$

and

$$\sum_{i=1}^{k} \tilde{\lambda}_i \mu_j^{n,i} > \tilde{\lambda}_j \mu_j^{n,j} - \sum_{i=1, i\neq j}^{k} \tilde{\lambda}_i |\mu_j^{n,i}| > \frac{\rho_0}{2k(1+\rho_0)} - \frac{\rho_0}{4k(1+\rho_0)} = \frac{\rho_0}{4k(1+\rho_0)} > 0,\ j \in I^*.$$

Finally, one completes the proof by noting that $\sum_{j\in I^*} \sum_{i=1}^{k} \tilde{\lambda}_i \mu_j^{n,i} = \sum_{i=1}^{k} \tilde{\lambda}_i \sum_{j \in I^*} \mu_j^{n,i} = 1.$ $\square$

As already said, Proposition 5 does not hold anymore for $\bar{d}_{n,p}$. As a consequence we have to establish a "firewall Lemma", which will be a useful tool to overcome this problem in the non-compact setting.

Lemma 2 (Firewall)**.** *Let $K \subset \mathbb{R}^d$ be compact and convex with $\mathring{K} \neq \emptyset$. Moreover, let $\varepsilon > 0$ be small enough so that*

$$K_\varepsilon = \{x \in K : \mathrm{dist}_{\ell^\infty}(x, K^c) \geq \varepsilon\} \neq \emptyset.$$

Let $\Gamma_{\alpha,\varepsilon}$ be a subset of the lattice $\alpha\mathbb{Z}^d$ with edge-length $\alpha > 0$ satisfying

$$K \setminus K_\varepsilon \subset \mathrm{conv}(\Gamma_{\alpha,\varepsilon}) \;\text{ and }\; \forall\, x \in K \setminus K_\varepsilon,\ \mathrm{dist}_{\|\cdot\|}(x, \Gamma_{\alpha,\varepsilon}) \leq C_{\|\cdot\|}\alpha$$

where $C_{\|\cdot\|} > 0$ is a real constant only depending on the norm $\|\cdot\|$.

Then, for every grid $\Gamma \subset \mathbb{R}^d$ containing K and every $\eta \in (0,1)$, it holds

$$\forall\, \xi \in K_\varepsilon, \quad F^p(\xi;\Gamma) \geq \frac{1}{(1+\eta)^{p+d+1}} F^p(\xi; (\Gamma \cap \mathring{K}) \cup \Gamma_{\alpha,\varepsilon}) - (1+\eta)^{-d-1}\eta^{-p}(d+1)\, C_{\|\cdot\|}^p \alpha^p.$$

Remark. The lattice $\Gamma_{\alpha,\varepsilon}$ and its size will be carefully defined and estimated for the specified compact sets K when calling upon the firewall lemma in what follows.

Proof. Let $\Gamma = \{x_1, \ldots, x_n\}$ and let $\xi \in K_\varepsilon$. Then we may choose $I = I(\xi) \subset \{1, \ldots, n\}$, $|I| \leq d+1$ such that

$$F^p(\xi;\Gamma) = \sum_{i \in I} \lambda_j \|\xi - x_i\|^p, \quad \sum_{i\in I} \lambda_i x_i = \xi,\ \lambda_i \geq 0, \sum_{i \in I} \lambda_i = 1.$$

If for every $x_i \in \Gamma \setminus \mathring{K}$ $\lambda_i = 0$ then $F^p(\xi, \Gamma) = F^p(\Gamma \cap \mathring{K})$ and our claim is trivial. Therefore, let $J(\xi) = \{i \,:\, x_i \in \Gamma \setminus \mathring{K},\, \lambda_i > 0\} \subset I(\xi)$ and choose one fixed $i_0 \in J(\xi)$. Let $\theta = \theta(i_0) \in (0,1)$ such that

$$\tilde{x}_{i_0} = \xi + \theta(x_{i_0} - \xi) \in K \setminus K_\varepsilon \quad \text{and} \quad \frac{\theta^{p\wedge 1}}{\theta + \lambda_{i_0}(1-\theta)} \le 1 + \eta$$

(when $p \ge 1$ the right constraint is empty). Setting

$$\tilde{\lambda}^0_i = \frac{\lambda_i \theta}{\theta + \lambda_{i_0}(1-\theta)},\ i \in I \setminus \{i_0\}, \quad \tilde{\lambda}^0_{i_0} = \frac{\lambda_{i_0}}{\theta + \lambda_{i_0}(1-\theta)}$$

we arrive at

$$\tilde{\lambda}^0_{i_0} \tilde{x}_{i_0} + \sum_{i \in I \setminus \{i_0\}} \tilde{\lambda}^0_i x_i = \xi,\ \tilde{\lambda}^0_i \ge 0,\ \sum_{i\in I} \tilde{\lambda}^0_i = 1.$$

Consequently

$$\begin{aligned}
\tilde{\lambda}^0_{i_0} \|\xi - \tilde{x}_{i_0}\|^p + \sum_{j \in I \setminus \{i_0\}} \tilde{\lambda}^0_i \|\xi - x_i\|^p &= \frac{\lambda_{i_0}\theta^p}{\theta + \lambda_{i_0}(1-\theta)} \|\xi - x_{i_0}\|^p + \sum_{i \in I \setminus \{i_0\}} \frac{\lambda_i \theta}{\theta + \lambda_{i_0}(1-\theta)} \|\xi - x_i\|^p \\
&\le \frac{\theta^{p\wedge 1}}{\theta + \lambda_{i_0}(1-\theta)} \sum_{i\in I} \lambda_i \|\xi - x_i\|^p \\
&\le (1+\eta) \sum_{i \in I} \lambda_i \|\xi - x_i\|^p.
\end{aligned}$$

Repeating the procedure for every $i \in J(\xi)$ finally yields by induction the existence of $\tilde{x}_i \in K \setminus K_\varepsilon$ and $\tilde{\lambda}_i$, $i \in I$ such that

$$\sum_{i \in I : x_i \notin \mathring{K}} \tilde{\lambda}_i \tilde{x}_i + \sum_{i \in I : x_i \in \mathring{K}} \tilde{\lambda}_i x_i = \xi,\ \tilde{\lambda}_i \ge 0,\ \sum_{i \in I} \tilde{\lambda}_i = 1$$

and

$$(1+\eta)^{|J(\xi)|} F^p(\xi; \Gamma) \ge \sum_{i \in I : x_i \notin \mathring{K}} \tilde{\lambda}_i \|\xi - \tilde{x}_i\|^p + \sum_{i \in I : x_i \in \mathring{K}} \tilde{\lambda}_i \|\xi - x_i\|^p. \tag{20}$$

Let us denote $\Gamma_{\alpha,\varepsilon} = \{a_1, \ldots, a_m\}$ and let $i_0 \in J(\xi)$ so that $\tilde{x}_{i_0}$ is a "modified" x_{i_0} (originally lying in $\Gamma \setminus \mathring{K}$). By construction $\tilde{x}_{i_0} \in K \setminus K_\varepsilon \subset \mathrm{conv}(\Gamma_{\alpha,\varepsilon})$ and there is $J_{i_0} \subset \{1, \ldots, m\}$ such that

$$F^p(\tilde{x}_{i_0}, \Gamma_{\alpha,\varepsilon}) = \sum_{j \in J_{i_0}} \mu^{i_0}_j \|\tilde{x}_{i_0} - a_j\|^p,\ \sum_{j \in J_{i_0}} \mu^{i_0}_j x_j = \tilde{x}_{i_0},\ \mu^{i_0}_j \ge 0,\ \sum_{j \in J_{i_0}} \mu^{i_0}_j = 1$$

and

$$\forall\, j \in J_{i_0}, \quad \|\tilde{x}_{i_0} - a_j\| \le C_{\|\cdot\|}\, \alpha.$$

Using the elementary inequality

$$\forall\, p > 0,\ \forall \eta > 0,\ \forall\, u, v \ge 0, \quad (u+v)^p \le (1+\eta)^p u^p + \Big(1 + \frac{1}{\eta}\Big)^p v^p,$$

we derive that for every $j \in J_{i_0}$

$$\|\xi - a_j\|^p \le \big(\|\xi - \tilde{x}_{i_0}\| + \|\tilde{x}_{i_0} - a_j\|\big)^p \le (1+\eta)^p \|\xi - \tilde{x}_{i_0}\|^p + \Big(1 + \frac{1}{\eta}\Big)^p C^p_{\|\cdot\|}\, \alpha^p.$$

As a consequence,

$$\sum_{j \in J_{i_0}} \mu^{i_0}_j \|\xi - a_j\|^p \le (1+\eta)^p \|\xi - \tilde{x}_{i_0}\|^p + \Big(1 + \frac{1}{\eta}\Big)^p C^p_{\|\cdot\|}\, \alpha^p$$

which in turn implies

$$\|\xi - \tilde{x}_{i_0}\|^p \geq \frac{1}{(1+\eta)^p} \sum_{j \in J_{i_0}} \mu_j^{i_0} \|\xi - a_j\|^p - \eta^{-p}\, C^p_{\|\cdot\|}\, \alpha^p.$$

Plugging this inequality in (20) yields and using that $|J(\xi)| \leq d+1$, we finally get

$$\begin{aligned}(1+\eta)^{|J(\xi)|} F^p(\xi;\Gamma) &\geq \sum_{i \in I : x_i \in \mathring{K}} \tilde{\lambda}_i \|\xi - x_i\|^p + \frac{1}{(1+\eta)^p} \sum_{i \in I : x_i \notin \mathring{K}} \tilde{\lambda}_i \sum_{j \in J_i} \mu_j^i \|\xi - a_j\|^p \\ &\quad - |J(\xi)| \eta^{-p} d\, C^p_{\|\cdot\|}\, \alpha^p \\ &\geq \frac{1}{(1+\eta)^p} F^p\big(\xi; (\Gamma \cap \mathring{K}\}) \cup \Gamma_{\alpha,\varepsilon}\big) - \eta^{-p}\,(d+1)\, C^p_{\|\cdot\|}\, \alpha^p. \square\end{aligned}$$

Now we can establish the sharp rate for the uniform distribution $U([0,1]^d)$.

Proposition 7 (Uniform distribution)**.** *For every $p \geq 1$,*

$$Q^{dq}_{\|\cdot\|,p,d} := \inf_{n \geq 0} n^{1/d}\, d_{n,p}\big(\mathcal{U}\big([0,1]^d\big)\big) = \lim_{n \to \infty} n^{1/d}\, d_{n,p}\big(\mathcal{U}\big([0,1]^d\big)\big).$$

Proof. Let $n, m \in \mathbb{N}$, $m < n$ and set $k = k(n,m) = \left\lfloor \left(\frac{n}{m}\right)^{1/d} \right\rfloor \geq 1$.

Covering the unit hypercube $[0,1]^d$ by k^d translates $C_1, \ldots, C_{k^d}$ of the hypercube $\left[0, \frac{1}{k}\right]^d$, we arrive at $\mathcal{U}\big([0,1]^d\big) = k^{-d} \sum_{i=1}^{k^d} \mathcal{U}(C_i)$. Hence, Proposition 5 yields

$$d^p_{n,p}\big(\mathcal{U}\big([0,1]^d\big)\big) \leq k^{-d} \sum_{i=1}^{k^d} d^p_m(\mathcal{U}(C_i)).$$

Furthermore, Proposition 6 implies

$$d_{m,p}(\mathcal{U}(C_i)) = k^{-1}\, d_{m,p}\big(\mathcal{U}\big([0,1]^d\big)\big),$$

so that we may conclude for all $n, m \in \mathbb{N}$, $m < n$,

$$d_{n,p}\big(\mathcal{U}\big([0,1]^d\big)\big) \leq k^{-1}\, d_{m,p}\big(\mathcal{U}\big([0,1]^d\big)\big).$$

Thus, we get

$$\begin{aligned} n^{1/d}\, d_{n,p}\big(\mathcal{U}\big([0,1]^d\big)\big) &\leq k^{-1}\, n^{1/d}\, d_{m,p}\big(\mathcal{U}\big([0,1]^d\big)\big) \\ &\leq \frac{k+1}{k}\, m^{1/d}\, d_{m,p}\big(\mathcal{U}\big([0,1]^d\big)\big),\end{aligned}$$

which yields for every fixed integer $m \geq 1$

$$\limsup_{n \to \infty} n^{1/d}\, d_{n,p}\big(\mathcal{U}\big([0,1]^d\big)\big) \leq m^{1/d}\, d_{m,p}\big(\mathcal{U}\big([0,1]^d\big)\big),$$

since $\lim_{n \to \infty} k(n,m) = +\infty$. This finally implies

$$\lim_{n \to \infty} n^{1/d}\, d_{n,p}\big(\mathcal{U}\big([0,1]^d\big)\big) = \inf_{m \geq 0} m^{1/d}\, d_{m,p}\big(\mathcal{U}\big([0,1]^d\big)\big). \qquad \square$$

Proposition 8. *For every $p \geq 1$,*

$$Q^{dq}_{\|\cdot\|,p,d} = \lim_{n \to \infty} n^{1/d}\, d_{n,p}\big(\mathcal{U}\big([0,1]^d\big)\big) = \lim_{n \to \infty} n^{1/d}\, \bar{d}_{n,p}\big(\mathcal{U}\big([0,1]^d\big)\big)$$

Proof. Since for every compactly supported distribution $\mathbf{P}$ we have $\bar{d}_{n,p}(\mathbf{P}) \leq d_{n,p}(\mathbf{P})$ it remains to show

$$Q^{\mathrm{dq}}_{\|\cdot\|,p,d} \leq \liminf_{n\to\infty} n^{1/d}\, \bar{d}_{n,p}\big(\mathcal{U}\big([0,1]^d\big)\big).$$

For $0 < \varepsilon < 1/2$ let $C_\varepsilon = (1/2, \dots, 1/2) + \frac{1-\varepsilon}{2}[-1,1]^d$ be the centered hypercube in $[0,1]^d$ with edge-length $1-\varepsilon$ and midpoint $(1/2, \dots, 1/2)$. Moreover let (Γ_n) be a sequence of quantizers such that, for every $n \geq 1$,

$$\bar{d}_p(\mathcal{U}\big([0,1]^d\big); \Gamma_n) \leq (1+\varepsilon)\bar{d}_{n,p}(\mathcal{U}\big([0,1]^d\big)).$$

Owing to Lemma 1, as $C_\varepsilon \subset (1,1)^d$, there is an integer $n_\varepsilon \in \mathbb{N}$ such that

$$\forall n \geq n_\varepsilon, \quad C_\varepsilon \subset \mathrm{conv}(\Gamma_n).$$

We therefore get for any $n \geq n_\varepsilon$

$$\begin{aligned}
(1+\varepsilon)^d \bar{d}^p_n\big(\mathcal{U}\big([0,1]^d\big)\big) &\geq \bar{d}^p\big(\mathcal{U}\big([0,1]^d\big); \Gamma_n\big) \\
&\geq \int_{C_\varepsilon} \bar{F}^p(\xi, \Gamma_n)^p d\xi = \int_{C_\varepsilon} F^p(\xi, \Gamma_n)^p d\xi = \lambda_d(C_\varepsilon) d^p\big(\mathcal{U}(C_\varepsilon), \Gamma_n\big) \\
&\geq (1-\varepsilon)^d d^p_n\big(\mathcal{U}(C_\varepsilon)\big) = (1-\varepsilon)^{d+p} d^p_n\big(\mathcal{U}\big([0,1]^d\big)\big)
\end{aligned}$$

where we used the scaling property (Proposition 6) in the last line.

Hence, we obtain for all $0 < \varepsilon < 1/2$

$$\liminf_{n\to\infty} n^{1/d}\, \bar{d}_{n,p}\big(\mathcal{U}\big([0,1]^d\big)\big) \geq \frac{(1-\varepsilon)^{1+d/p}}{(1+\varepsilon)^{d/p}}\, Q^{\mathrm{dq}}_{\|\cdot\|,p,d},$$

so that letting $\varepsilon \to 0$ completes the proof. □

Proposition 9. *Let* $\mathbf{P} = \sum_{i=1}^m s_i \mathcal{U}(C_i)$, $\sum_{i=1}^m s_i = 1$, $s_i > 0$, $i = 1, \dots, m$, *where* $C_i = a_i + [0,l]^d$, $i = 1, \dots, m$, *are pairwise disjoint hypercubes in* $\mathbb{R}^d$ *with common edge-length* l. *Set*

$$h := \frac{d\mathbf{P}}{d\lambda_d} = \sum_{i=1}^m s_i l^{-d} \mathbb{1}_{C_i}.$$

Then

$$\lim_{n\to\infty} n^{1/d}\, d_{n,p}(\mathbf{P}) = \lim_{n\to\infty} n^{1/d}\, \bar{d}_{n,p}(\mathbf{P}) = Q^{dq}_{\|\cdot\|,p,d} \cdot \|h\|^{\frac{1}{p}}_{d/(d+p)}.$$

Proof. Since $d_{n,p}(\mathbf{P}) \geq \bar{d}_{n,p}(\mathbf{P})$ it suffices to show that

$$\limsup_{n\to\infty} n^{1/d}\, d_{n,p}(\mathbf{P}) \leq Q^{\mathrm{dq}}_{\|\cdot\|,p,d} \cdot \|h\|^{\frac{1}{p}}_{d/(d+p)} \quad \text{and} \quad \liminf_{n\to\infty} n^{1/d}\, \bar{d}_{n,p}(\mathbf{P}) \geq Q^{\mathrm{dq}}_{\|\cdot\|,p,d} \cdot \|h\|^{\frac{1}{p}}_{d/(d+p)}.$$

For $n \in \mathbb{N}$, set

$$t_i = \frac{s_i^{d/(d+p)}}{\sum_{j=1}^m s_j^{d/(d+p)}} \qquad \text{and} \qquad n_i = \lfloor t_i n \rfloor,\ 1 \leq i \leq m.$$

Then, by Proposition 5 and Proposition 6, we get for every $n \geq \max_{1\leq i\leq m}(1/t_i)$

$$d^p_n(\mathbf{P}) \leq \sum_{i=1}^m s_i\, d^p_n(\mathcal{U}(C_i)) = l^p \sum_{i=1}^m s_i\, d^p_{n_i}(\mathcal{U}\big([0,1]^d\big)).$$

Proposition 7 then yields

$$n^{\frac{p}{d}}\, d^p_{n_i}(\mathcal{U}\big([0,1]^d\big)) = \left(\frac{n}{n_i}\right)^{\frac{p}{d}} n_i^{\frac{p}{d}}\, d^p_{n_i}(\mathcal{U}\big([0,1]^d\big)) \longrightarrow t_i^{-\frac{p}{d}} Q^{\mathrm{dq}}_{\|\cdot\|,p,d} \quad \text{as } n \to +\infty.$$

Noting that $\|h\|_{d/(d+p)} = l^p \Big(\sum s_i^{d/(d+p)}\Big)^{(d+p)/d}$, we get

$$\limsup_{n\to\infty} n^{\frac{p}{d}}\, d^p_{n,p}(\mathbf{P}) \le Q^{\mathrm{dq}}_{\|\cdot\|,p,d} l^p \sum_{i=1}^m s_i\, t_i^{-\frac{p}{d}} = Q^{\mathrm{dq}}_{\|\cdot\|,p,d} \cdot \|h\|_{d/(d+p)}.$$

(b) Let $\varepsilon \in (0, l/2)$ and let $C_{i,\varepsilon}$ denote the closed hypercube with the same center as C_i but with edge-length $l-\varepsilon$. For $\alpha \in (0,\varepsilon/2)$, we set $\tilde\alpha = \frac{l}{\lceil l/\alpha\rceil}$ and we define the lattice

$$\Gamma_{\alpha,\varepsilon,i} = \big(a_i + \tilde\alpha\mathbb{Z}^d\big) \cap (C_i \setminus C_{i,\varepsilon}) \bigcup \{\text{vertices of } C_i\}.$$

It is clear that $\mathrm{conv}(\Gamma_{\alpha,\varepsilon,i}) = C_i \subset C_i \setminus C_{i,\varepsilon}$ since it contains the vertices of C_i. Moreover, for every $\xi \in C_i \setminus C_{i,\varepsilon}$, $\mathrm{dist}_{\ell^\infty}(\xi, \Gamma_{\alpha,\varepsilon,i}) \le \alpha$ so that there exists a real constant $C_{\|\cdot\|} > 0$ only depending on the norm $\|.\|$ such that $\mathrm{dist}_{\|.\|}(\xi,\Gamma_{\alpha,\varepsilon,i}) \le C_{\|\cdot\|}\alpha$. Consequently the lattice $\Gamma_{\alpha,\varepsilon,i}$ satisfies the assumption of the firewall lemma (Lemma 2).

On the other hand, easy combinatorial arguments show that number of points m_i of $\Gamma_{\alpha,\varepsilon,i}$ falling in C_i satisfies $\lceil \frac{l}{\tilde\alpha}\rceil^d \le m_i \le \big(\lceil\frac{l}{\tilde\alpha}\rceil + 1\big)^d + 2^d$ whereas the number $m_{i,\varepsilon}$ of points falling in $C_{i,\varepsilon}$ satisfies $\big(\lceil\frac{l-\varepsilon}{\tilde\alpha}\rceil - 1\big)^d \le m_{i,\varepsilon} \le \big(\lceil\frac{l-\varepsilon}{\tilde\alpha}\rceil + 1\big)^d$ so that

$$\left\lceil\frac{l}{\tilde\alpha}\right\rceil^d - \left(\left\lceil\frac{l-\varepsilon}{\tilde\alpha}\right\rceil + 1\right)^d \le |\Gamma_{\alpha,\varepsilon,i}| \le \left(\left\lceil\frac{l}{\tilde\alpha}\right\rceil + 1\right)^d + 2^d - \left(\left\lceil\frac{l-\varepsilon}{\tilde\alpha}\right\rceil - 1\right)^d.$$

We define for every $\varepsilon \in (0,l/2), \alpha \in (0,\varepsilon/2)$

$$g_{l,\varepsilon}(\alpha) = \alpha^d |\Gamma_{\alpha,\varepsilon,i}|.$$

Since $\frac{\alpha}{\tilde\alpha} \to 1$ and $2\alpha\left\lceil\frac{\varepsilon/2}{\tilde\alpha}\right\rceil \to \varepsilon$ as $\alpha \to 0$, we conclude from the above inequalities that

$$\forall\varepsilon \in (0,l/2), \quad \lim_{\alpha\to 0} g_{l,\varepsilon}(\alpha) = l^d - (l-\varepsilon)^d. \tag{21}$$

Let $\eta \in (0,1)$ and denote by Γ_n a sequence of n-quantizers such that $\bar d^p(\mathbf{P};\Gamma_n) \le (1+\eta) d^p_n(\mathbf{P})$. It follows from Proposition 2 that $\bar d^p(\mathbf{P};\Gamma_n) \to 0$ for $n \to \infty$ so that Lemma 1 yields the existence of $n_\varepsilon \in \mathbb{N}$ such that for any $n \ge n_\varepsilon$

$$\bigcup_{1\le i\le m} C_{i,\varepsilon} \subset \mathrm{conv}(\Gamma_n).$$

We then derive from Lemma 2 (firewall)

$$\begin{aligned}
\bar d^p(\mathcal{U}(C_i);\Gamma_n) &= l^{-d}\int_{C_i} \bar F^p(\xi;\Gamma_n)\,\lambda_d(d\xi)\\
&\ge l^{-d}\int_{C_{i,\varepsilon}} \bar F^p(\xi;\Gamma_n)\,\lambda_d(d\xi) = l^{-d}\int_{C_{i,\varepsilon}} F^p(\xi;\Gamma_n)\,\lambda_d(d\xi)\\
&\ge \frac{l^{-d}\,(l-\varepsilon)^d}{(1+\eta)^{p+d+1}}\, d^p\big(\mathcal{U}(C_{i,\varepsilon});(\Gamma_n\cap \mathring{C}_i)\cup\Gamma_{\alpha,\varepsilon,i}\big) - l^{-d}\,(l-\varepsilon)^d\frac{(1+\eta)^{-d-1}}{\eta^p}\,(d+1)C_{\|\cdot\|}\cdot\alpha^p.
\end{aligned}$$

At this stage, we set for every $\rho > 0$

$$\alpha_n = \alpha_n(\rho) = \Big(\frac{m}{\rho n}\Big)^{1/d} \tag{22}$$

and denote

$$n_i = |(\Gamma_n \cap \mathring{C}_i)\cup\Gamma_{\alpha_n,\varepsilon,i}|.$$

Proposition 6 yields $d_{n_i,p}(\mathcal{U}(C_{i,\varepsilon})) = (l-\varepsilon) d_{n_i,p}(\mathcal{U}\big([0,1]^d\big))$, so that we get

$$
\begin{aligned}
n^{\frac{p}{d}} d_n^p(\mathbf{P}) &\geq \frac{1}{1+\eta} \sum_{i=1}^m s_i\, n^{\frac{p}{d}}\, \bar{d}^p(\mathcal{U}(C_i); \Gamma_n) \\
&\geq \frac{l^{-d}\,(l-\varepsilon)^d}{(1+\eta)^{p+d+2}} \sum_{i=1}^m s_i\, n^{\frac{p}{d}}\, d^p\big(\mathcal{U}(C_{i,\varepsilon}); (\Gamma_n \cap \mathring{C}_i) \cup \Gamma_{\alpha_n,\varepsilon,i}\big) \\
&\qquad - l^{-d}\,(l-\varepsilon)^d \frac{(1+\eta)^{-d-2}}{\eta^p} \sum_{i=1}^m s_i\,(d+1)\, C_{\|\cdot\|} \cdot \alpha^p \cdot n^{\frac{p}{d}} \\
&\geq \frac{l^{-d}\,(l-\varepsilon)^{d+p}}{(1+\eta)^{p+d+2}} \sum_{i=1}^m s_i\, n^{\frac{p}{d}}\, d_{n_i}^p\big(\mathcal{U}\big([0,1]^d\big)\big) - l^{-d}\,(l-\varepsilon)^d \frac{(1+\eta)^{-d-2}}{\eta^p} (d+1)\, C_{\|\cdot\|} \Big(\frac{m}{\rho}\Big)^{\frac{p}{d}}.
\end{aligned}
\tag{23}
$$

Since

$$
\frac{n_i}{n} \leq \frac{|\Gamma_n \cap \mathring{C}_i|}{n} + \frac{g_{l,\varepsilon}(\alpha_n)}{n\alpha_n^d} = \frac{|\Gamma_n \cap \mathring{C}_i|}{n} + \frac{\rho}{m} g_{l,\varepsilon}(\alpha_n),
$$

we conclude from (21) and (22) that

$$
\limsup_{n\to\infty} \sum_{i=1}^m \frac{n_i}{n} \leq 1 + \rho\big(l^d - (l-\varepsilon)^d\big).
$$

We may choose a subsequence (still denoted by (n)), such that

$$
n^{1/d}\, \bar{d}_{n,p}(\mathbf{P}) \to \liminf_{n\to\infty} n^{1/d}\, d_{n,p}(\mathbf{P}) \qquad \text{and} \qquad \frac{n_i}{n} \to v_i \in [0, 1 + \rho(l^d - (l-\varepsilon)^d)].
$$

As a matter of fact, $v_i > 0$, for every $i = 1, \ldots m$: otherwise Proposition 7 would yield

$$
\begin{aligned}
n^{\frac{p}{d}}\, \bar{d}_{n,p}^p(\mathbf{P}) \geq & \frac{l^{-d}\,(l-\varepsilon)^{d+p}}{(1+\eta)^{p+d+2}} \sum_{i=1}^m s_i \left(\frac{n_i}{n}\right)^{-\frac{p}{d}} n_i^{\frac{p}{d}}\, d_{n_i}^p\big(\mathcal{U}\big([0,1]^d\big)\big) \\
& - l^{-d}\,(l-\varepsilon)^d \frac{(1+\eta)^{p-d-2}}{\eta^p} (d+1) C_{\|\cdot\|} \cdot \left(\frac{m}{\rho}\right)^{\frac{p}{d}} \\
& \to +\infty \quad \text{as } n \to +\infty
\end{aligned}
$$

which contradicts (a). Consequently, we may normalize the v_i's by setting

$$
\widetilde{v}_i = \frac{v_i}{1 + \rho(l^d - (l-\varepsilon)^d)},\ i = 1, \ldots, m,
$$

so that $\sum_{i=1}^m \widetilde{v}_i \leq 1$. We derive from Proposition 7 that

$$
\begin{aligned}
\liminf_{n\to\infty} \sum_{i=1}^m s_i\, n^{\frac{p}{d}}\, d_{n_i}^p\big(\mathcal{U}\big([0,1]^d\big)\big) &\geq \sum_{i=1}^m s_i\, v_i^{-\frac{p}{d}} n_i^{\frac{p}{d}}\, d_{n_i}^p\big(\mathcal{U}\big([0,1]^d\big)\big) \\
&= Q^{\rm dq}_{\|\cdot\|,p,d} (1 + \rho(l^d - (l-\varepsilon)^d)^{-\frac{p}{d}} \sum_{i=1}^m s_i\, \widetilde{v}_i^{-\frac{p}{d}} \\
&\geq Q^{\rm dq}_{\|\cdot\|,p,d} (1 + \rho(l^d - (l-\varepsilon)^d)^{-\frac{p}{d}} \inf_{\sum_i y_i \leq 1, y_i \geq 0} \sum_{i=1}^m s_i y_i^{-\frac{p}{d}} \\
&= Q^{\rm dq}_{\|\cdot\|,p,d} (1 + \rho(l^d - (l-\varepsilon)^d)^{-\frac{p}{d}} \left(\sum_{i=1}^m s_i^{d/(d+p)}\right)^{(d+p)/d}.
\end{aligned}
$$

Hence, we derive from (23)

$$\liminf_{n\to\infty} n^{\frac{p}{d}}\,\bar d^{p}_{n,p}(\mathbf{P}) \geq \frac{l^{-d}\,(l-\varepsilon)^{d+p}}{(1+\eta)^{p+d+2}\big(1+\rho(l^d-(l-\varepsilon)^d)\big)^{\frac{p}{d}}}\, Q^{\mathrm{dq}}_{\|\cdot\|,p,d}\Big(\sum_{i=1}^{m} s_i^{d/(d+p)}\Big)^{(d+p)/d} - l^{-d}\,(l-\varepsilon)^d\,\frac{(1+\eta)^{-d-2}}{\eta^p}(d+1)\,C_{\|\cdot\|}\cdot\Big(\frac{m}{\rho}\Big)^{\frac{p}{d}}.$$

Letting $\varepsilon\to 0$ implies

$$\begin{aligned}\liminf_{n\to\infty} n^{\frac{p}{d}}\,\bar d^{p}_{n,p}(\mathbf{P}) &\geq \frac{l^p}{(1+\eta)^{p+d+2}}\, Q^{\mathrm{dq}}_{\|\cdot\|,p,d}\Big(\sum_{i=1}^{m} s_i^{d/(d+p)}\Big)^{(d+p)/d} - \frac{(1+\eta)^{-d-2}}{\eta^p}(d+1)\,C_{\|\cdot\|}\Big(\frac{m}{\rho}\Big)^{\frac{p}{d}}\\ &= \frac{1}{(1+\eta)^{p+d+2}}\, Q^{\mathrm{dq}}_{\|\cdot\|,p,d}\cdot\|h\|_{d/(d+p)} - \frac{(1+\eta)^{-d-2}}{\eta^p}d\,C_{\|\cdot\|}\Big(\frac{m}{\rho}\Big)^{\frac{p}{d}}\end{aligned}$$

and, finally, letting successively ρ go to $+\infty$ and η go to 0 completes the proof. □

Proposition 10. *Assume that* $\mathbf{P}$ *is absolutely continuous w.r.t.* λ_d *with compact support. Then*

$$\lim_{n\to\infty} n^{\frac{p}{d}}\,d_{n,p}(\mathbf{P}) = \liminf_{n\to\infty} n^{\frac{p}{d}}\,\bar d_{n,p}(\mathbf{P}) = Q^{dq}_{\|\cdot\|,p,d}\cdot\|h\|^{\frac{1}{p}}_{d/(d+p)}$$

Proof. Since $d_{n,p}(\mathbf{P})\geq \bar d_{n,p}(\mathbf{P})$ it suffices to show that

$$\limsup_{n\to\infty} n^{\frac{p}{d}}\,d_{n,p}(\mathbf{P}) \leq Q^{\mathrm{dq}}_{\|\cdot\|,p,d}\cdot\|h\|^{\frac{1}{p}}_{d/(d+p)} \quad\text{and}\quad \liminf_{n\to\infty} n^{\frac{p}{d}}\,\bar d_{n,p}(\mathbf{P}) \geq Q^{\mathrm{dq}}_{\|\cdot\|,p,d}\cdot\|h\|^{\frac{1}{p}}_{d/(d+p)}.$$

Preliminary step. Let $C=[-l/2,l/2]^d$ be a closed hyper hypercube centered at the origin, parallel to the coordinate axis with edge-length l, such that $\mathrm{supp}(\mathbf{P})\subset C$. For $k\in\mathbb{N}$ consider the tessellation of C into k^d closed hypercubes with common edge-length l/k. To be precise, for every $\underline{i}=(i_1,\ldots,i_d)\in\mathbb{Z}^d$, we set

$$C_{\underline{i}} = \prod_{r=1}^{d}\Big[-\frac{l}{2}+\frac{i_r l}{k}, -\frac{l}{2}+\frac{(i_r+1)l}{k}\Big].$$

Then, set

$$h=\frac{d\mathbf{P}}{d\lambda_d}\ \text{ and }\ \mathbf{P}_k=\sum_{\substack{\underline{i}\in\mathbb{Z}^d\\ 0\leq i_r<k}}\mathbf{P}(C_{\underline{i}})\,\mathcal{U}(C_{\underline{i}}),\ h_k=\frac{d\mathbf{P}_k}{d\lambda_d}=\sum_{\substack{\underline{i}\in\mathbb{Z}^d\\ 0\leq i_r<k}}\frac{\mathbf{P}(C_{\underline{i}})}{\lambda_d(C_{\underline{i}})}\mathbb{1}_{C_{\underline{i}}},\ k\geq 1. \tag{24}$$

By differentiation of measures we obtain $h_k\to h$, λ_d-*a.s.* as $k\to\infty$. Which in turn implies, owing to Scheffé's Lemma,

$$\lim_{k\to+\infty}\|h_k-h\|_1=0.$$

Furthermore,

$$\lim_{k\to+\infty}\|h_k\|_{d/(d+p)}=\|h\|_{d/(d+p)}$$

since $\|h_k-h\|_{d/(d+p)}\leq \Big(\lambda_d(C)\Big)^{\frac{p}{d}}\|h_k-h\|_1$ by Jensen's Inequality applied to the probability measure $\frac{\lambda_{d\,|C}}{\lambda_d(C)}$. Moreover, by Proposition 9 we have

$$\lim_{n\to\infty} n^{1/d}\,d_{n,p}(\mathbf{P}_k) = Q^{\mathrm{dq}}_{\|\cdot\|,p,d}\,\|h_k\|^{\frac{1}{p}}_{d/(d+p)}. \tag{25}$$

Likewise, we define an inner approximation of $\mathbf{P}$: denote by

$$C^k = \bigcup_{C_{\underline{i}} \subset \overbrace{\operatorname{supp}(\mathbf{P})}^{\circ}} C_{\underline{i}}$$

the union of the hypercubes $C_{\underline{i}}$ lying in the interior of $\operatorname{supp}(\mathbf{P})$. Setting

$$\mathring{\mathbf{P}}_k = \sum_{C_{\underline{i}} \subset \overbrace{\operatorname{supp}(\mathbf{P})}^{\circ}} \mathbf{P}(C_{\underline{i}})\,\mathcal{U}(C_{\underline{i}}) \quad \text{and} \quad \mathring{h}_k \;=\frac{d\mathring{\mathbf{P}}_k}{d\lambda_d} = h_k \mathbb{1}_{C^k},$$

we have as above that

$$\mathring{h}_k \to h, \quad \lambda_d\text{-a.s.} \quad \text{as } k \to +\infty.$$

Consequently we also have

$$\lim_{k\to\infty} \|\mathring{h}_k - h\|_1 = 0 \quad \text{and} \quad \lim_{k\to\infty} \|\mathring{h}_k\|_{d/(d+p)} = \|h\|_{d/(d+p)}.$$

We get likewise by Proposition 9 that, for every $k \in \mathbb{N}$,

$$\lim_{n\to\infty} n^{1/d}\, d_{n,p}(\mathring{\mathbf{P}}_k) = Q^{\mathrm{dq}}_{\|\cdot\|,p,d} \cdot \|\mathring{h}_k\|^{\frac{1}{p}}_{d/(d+p)}. \tag{26}$$

(a) Let $0 < \varepsilon < 1$ and $n \geq 2^d/\varepsilon$. If we divide each edge of the hypercube C into

$$m = \left\lfloor (\varepsilon n)^{1/d} \right\rfloor - 1$$

intervals of equal length l/m, the interval endpoints define $m+1$ grid points on each edge. Denoting by $\Gamma_1 = \Gamma_1(\varepsilon, n)$ the product quantizer made up by this procedure, we clearly have

$$|\Gamma_1| = (m+1)^d = \left\lfloor (\varepsilon n)^{1/d} \right\rfloor^d =: n_1.$$

For this product quantizer it follows from Proposition 4 that, for all $\xi \in C$,

$$F^p(\xi;\Gamma_1) \leq C_{\|\cdot\|} \sum_{i=1}^{d} \left(\frac{l}{2m}\right)^p \leq C_{\|\cdot\|,p,d}\, \frac{l^p}{(\varepsilon n)^{\frac{p}{d}}}.$$

For $n_2 = \lfloor (1-\varepsilon) n \rfloor$ let Γ_2 be an n_2-quantizer such that $d^p(\mathbf{P}_k;\Gamma_2) \leq (1+\varepsilon) d^p_{n_2}(\mathbf{P}_k)$. We clearly have $|\Gamma_1 \cup \Gamma_2| \leq n$ and

$$\begin{aligned} n^{\frac{p}{d}} \left| \int F^p(\xi;\Gamma_1\cup\Gamma_2) d\mathbf{P}_k(\xi) - \int F^p(\xi;\Gamma_1\cup\Gamma_2) d\mathbf{P}(\xi) \right| &\leq n^{\frac{p}{d}} \int F^p(\xi;\Gamma_1\cup\Gamma_2) |h_k(\xi) - h(\xi)| d\lambda_d \xi \\ &\leq C_{\|\cdot\|,p,d}\, \frac{l^p}{\varepsilon^{\frac{p}{d}}} \|h_k - h\|_1 = c_{1,\varepsilon} \|h_k - h\|_1 \end{aligned}$$

for $k \in \mathbb{N}$ and $n \geq \max\left\{\frac{2^d}{\epsilon}, \frac{1}{1-\varepsilon}\right\}$. This implies

$$\begin{aligned} n^{\frac{p}{d}} d^p_n(\mathbf{P}) &\leq n^{\frac{p}{d}} \int F^p(\xi;\Gamma_1\cup\Gamma_2) d\mathbf{P}(\xi) \\ &\leq n^{\frac{p}{d}} \int F^p(\xi;\Gamma_1\cup\Gamma_2) d\mathbf{P}_k(\xi) + c_1 \|h_k - h\|_1 \\ &\leq n^{\frac{p}{d}} \int F^p(\xi;\Gamma_2) d\mathbf{P}_k(\xi) + c_1 \|h_k - h\|_1 \\ &\leq (1+\varepsilon)\, n^{\frac{p}{d}} d^p_{n_2}(\mathbf{P}_k) + c_{1,\varepsilon} \|h_k - h\|_1, \end{aligned}$$

so that we can conclude from (25) that

$$\limsup_{n\to\infty} n^{\frac{p}{d}} d_n^p(\mathbf{P}) \le \frac{1+\varepsilon}{(1-\varepsilon)^{\frac{p}{d}}} (Q^{\mathrm{dq}}_{\|\cdot\|,p,d})^p \|h_k\|_{d/(d+p)} + c_{1,\varepsilon}\|h_k - h\|_1.$$

Letting first k go to infinity and then letting ε go to zero yields

$$\limsup_{n\to\infty} n^{1/d} d_n^p(\mathbf{P}) \le Q^{\mathrm{dq}}_{\|\cdot\|,p,d}\|h_k\|^{\frac{1}{p}}_{d/(d+p)}.$$

(b) Assume now that Γ_3 is an n_2-quantizer such that $\bar{d}^p(\mathbf{P};\Gamma_3) \le (1+\varepsilon)\,\bar{d}^p_{n_2}(\mathbf{P})$. Again it holds $|\Gamma_1 \cup \Gamma_3| \le n$ and we derive as above

$$n^{\frac{p}{d}}\left|\int F^p(\xi;\Gamma_1\cup\Gamma_3)d\mathring{\mathbf{P}}_k(\xi) - \int F^p(\xi;\Gamma_1\cup\Gamma_3)d\mathbf{P}(\xi)\right| \le c_{2,\varepsilon}\|\mathring{h}_k - h\|_1. \tag{27}$$

Moreover, Lemma 1 yields for every $k \in \mathbb{N}$ the existence of $n_{k,\varepsilon} \in \mathbb{N}$ such that, for all $n \ge n_{k,\varepsilon}$,

$$\begin{aligned}(1+\varepsilon)\,\bar{d}^p_{n_2}(\mathbf{P}) \ge \bar{d}^p(\mathbf{P};\Gamma_3) &\ge \int_{\mathrm{conv}(\Gamma_3)} F^p(\xi;\Gamma_3)d\mathbf{P}(\xi)\\ &\ge \int_{C^k} F^p(\xi;\Gamma_3)d\mathbf{P}(\xi) \ge \int_{C^k} F^p(\xi;\Gamma_1\cup\Gamma_3)d\mathbf{P}(\xi).\end{aligned}$$

Thus, we derive from (27) that, for every $n \ge \max\left(n_{k,\varepsilon}, \frac{2^d}{\varepsilon}, \frac{1}{1-\varepsilon}\right)$,

$$\begin{aligned}(1+\varepsilon)\,n^{\frac{p}{d}}\,\bar{d}^p_{n_2}(\mathbf{P}) &\ge n^{\frac{p}{d}}\int_{C^k} F^p(\xi;\Gamma_1\cup\Gamma_3)d\mathbf{P}(\xi)\\ &\ge n^{\frac{p}{d}}\int_{C^k} F^p(\xi;\Gamma_1\cup\Gamma_3)d\mathring{\mathbf{P}}_k(\xi) - c_{2,\varepsilon}\|\mathring{h}_k - h\|_1\\ &\ge n^{\frac{p}{d}} d^p_n(\mathring{\mathbf{P}}_k) - c_{2,\varepsilon}\|\mathring{h}_k - h\|_1,\end{aligned}$$

which yields, once combined with (26),

$$\frac{1+\varepsilon}{(1-\varepsilon)^{\frac{p}{d}}}\liminf_{n\to\infty} n_2^{\frac{p}{d}}\,\bar{d}^p_{n_2,p}(\mathbf{P}) \ge Q^{\mathrm{dq}}_{\|\cdot\|,p,d}\|\mathring{h}_k\|_{d/(d+p)} - c_{2,\varepsilon}\|\mathring{h}_k - h\|_1.$$

Letting first k go to ∞ and then letting ε go to 0, we get

$$\liminf_{n\to\infty} n^{\frac{1}{d}}\,\bar{d}_{n,p}(\mathbf{P}) \ge Q^{\mathrm{dq}}_{\|\cdot\|,p,d}\|h\|^{\frac{1}{p}}_{d/(d+p)}. \square$$

Proposition 11 (Singular distribution)**.** *Assume that* $\mathbf{P}$ *is singular with respect to* λ_d *and has compact support. Then*

$$\limsup_{n\to\infty} n^{\frac{p}{d}}\,\bar{d}_{n,p}(\mathbf{P}) = 0.$$

Proof. Let A be a Borel set such that $\mathbf{P}(A) = 1$ and $\lambda_d(A) = 0$. Let $\varepsilon > 0$; by the outside regularity of λ_d, there exists an open set $O = O(\varepsilon) \supset A$ such that $\lambda_d(O) \le \varepsilon$ (and $\mathbf{P}(O) = 1$). Let C be an open hypercube with edges parallel to the coordinate axis, edge-length ℓ and containing the closure of A.

Let $C_k = \prod_{i=1}^d [c_{k,i}, c_{k,i} + \ell_i)$, $k \in \mathbb{N}$, be a countable partition of O consisting of nonempty half-open hypercubes, still with edges parallel to the coordinate axis (see, *e.g.* Lemma 1.4.2 in [4]).

Let $m = m(\varepsilon) \in \mathbb{N}$ such that $\displaystyle\sum_{k\ge m+1} \mathbf{P}(C_k) \le \varepsilon^{\frac{p}{d}}\ell^{-p}$.

Let $n\in\mathbb{N}$, $n\geq 2^{d+1}$ and let $n_1,\ldots,n_d\geq 2$ be integers such that the product $n_1^d+\cdots+n_m^d\leq n/2$. One designs a grid Γ as follows.

For every $k\in\{1,\ldots,m\}$, we consider the lattice of C_k of size n_i^d defined by

$$\prod_{i=1}^{d}\Big\{c_{k,i}+\frac{r_i}{n_k-1}\ell_i,\; r_i=0,\ldots,n_k-1,\; i=1,\ldots,d\Big\}.$$

Then, one defines likewise the lattice of C of size $n_{m+1}^d\leq n/2$

$$\prod_{i=1}^{d}\Big\{c_{k,i}+\frac{r_i}{n_{m+1}-1}\ell_i,\; r_i=0,\ldots,n_{m+1}-1,\; i=1,\ldots,d\Big\}.$$

The grid Γ is made up with all the points of the $m+1$ above finite lattices.

Now let $\xi\in A$. It is clear from the definition of the function F_p that

$$F_p(\xi;\Gamma)\leq\left\{\begin{array}{ll} C_{\|.\|}\big(\ell_k/n_k\big)^p & \text{if } \xi\in\bigcup_{k=1}^m C_k\\ C_{\|.\|}\big(\ell/n_{m+1}\big)^p & \text{if } \xi\in C\setminus\bigcup_{k=1}^m C_k\end{array}\right.$$

where $C_{\|.\|}>0$ is a real constant only depending on the norm. As a consequence

$$\begin{aligned} d_n^p(\mathbf{P}) &= \sum_{k=1}^m\int_{C_k}F^p(\xi;\Gamma)d\mathbf{P}(\xi)+\int_{C\setminus\bigcup_{k=1}^m C_k}F^p(\xi;\Gamma)d\mathbf{P}(\xi)\\ &\leq C_{\|.\|}\Big(\sum_{k=1}^m(\ell_k/n_k)^p\mathbf{P}(C_k)+(\ell/n_{m+1})^p\mathbf{P}(C\setminus\bigcup_{k=1}^m C_k)\Big).\end{aligned}$$

Set for every $k\in\{1,\ldots,m\}$, $n_k=\left\lfloor\dfrac{\ell_k(n/2)^{\frac1d}}{(\sum_{k'=1}^d\ell_{k'}^d)^{\frac1d}}\right\rfloor$ and $n_{m+1}=\lfloor(n/2)^{\frac1d}\rfloor$. Note that

$$\sum_{k'=1}^d\ell_{k'}^d=\sum_{k=1}^m\lambda_d(C_k)\leq\lambda_d(O)\leq\varepsilon.$$

Elementary computations show that for large enough n, all the integers n_k are greater than 1 and that

$$\begin{aligned}\sum_{k=1}^m(\ell_k/n_k)^p\mathbf{P}(C_k)+(\ell/n_{m+1})^p\mathbf{P}(C\setminus\bigcup_{k=1}^m C_k) &\leq (\sum_{k'=1}^d\ell_{k'}^d)^{\frac pd}(n/2)^{-\frac pd}\mathbf{P}\big(\cup_{1\leq k\leq m}C_k\big)+\\ &\quad +(n/2)^{-\frac pd}\ell^p\mathbf{P}\big(C\setminus\bigcup_{k=1}^m C_k\big)\end{aligned}$$

so that

$$\limsup_n n^{\frac pd}d_n^p(\mathbf{P})\leq C_{\|.\|}(\varepsilon/2)^{\frac pd}$$

which in turn implies, by letting ε go to 0, that $\limsup_n n^{\frac pd}d_n^p(\mathbf{P})=0$. $\square$

Proof of Theorem 2: Claim (a) follows directly from Propositions 10 , 11 and Proposition 5: Assume $\mathbf{P}=\rho\mathbf{P}_a+(1-\rho)\mathbf{P}_s$ where $\mathbf{P}_a=\frac{h}{\rho}\lambda_d$ and $\mathbf{P}_s$ denote the absolutely continuous and singular part of $\mathbf{P}$ respectively. The following inequalities hold true

$$\rho\bar{d}_{n,p}(\mathbf{P}_a)\leq\bar{d}_{n,p}(\mathbf{P})\leq\rho\bar{d}_{n_1,p}(\mathbf{P}_a)+(1-\rho)\bar{d}_{n_2,p}(\mathbf{P}_s)$$

for every triplet of integers (n_1,n_2,n) with $n_1+n_2\le n$. Set $n_1=n_1(n)=\lfloor(1-\varepsilon n)\rfloor$, $n_2=n_2(n)=\lfloor\varepsilon n\rfloor$. Then we derive that

$$\rho Q^{\rm dq}_{\|\cdot\|,p,d}\cdot\|\frac{h}{\rho}\|^{\frac1p}_{d/(d+p)}\liminf_n n^{\frac pd}\bar d_{n,p}(\mathbf{P}_a)\le\liminf_n n^{\frac pd}\bar d_{n,p}(\mathbf{P})\le\limsup_n n^{\frac pd}\bar d_{n,p}(\mathbf{P})\le\rho(1-\varepsilon)^{-\frac pd}Q^{\rm dq}_{\|\cdot\|,p,d}\cdot\|\frac{h}{\rho}\|^{\frac1p}_{d/(d+p)}$$

Letting ε go to 0 completes the proof.

Furthermore, part (c) was derived in [12], Section 5.1. Hence, it remains to prove (b)

Proof. Step 1. (Lower bound) If X is compactly supported, the assertion follows from Proposition 10. Otherwise, set for every $R\in(0,\infty)$,

$$C_{_R}=[-R,R]^d\ \text{ and }\ \mathbf{P}(\cdot|C_k)=\frac{h\mathbb{1}_{C_k}}{\mathbf{P}(C_k)}\lambda_d,\ k\in\mathbb{N}.$$

Proposition 10 yields again

$$\lim_{n\to\infty}n^{\frac1d}\,\bar d_{n,p}(\mathbf{P}(\cdot|C_k))=Q^{\rm dq}_{\|\cdot\|,p,d}\cdot\|h\mathbb{1}_{C_k}/\mathbf{P}(C_k)\|^{\frac1p}_{d/(d+p)},\tag{28}$$

so that $\bar d^p_{n,p}(\mathbf{P})\ge\mathbf{P}(\cdot|C_k)\bar d^p_{n,p}(\mathbf{P}(\cdot|C_k))$ implies for all $k\in\mathbb{N}$

$$\liminf_{n\to\infty}n^{\frac1d}\,\bar d_{n,p}(\mathbf{P})\ge Q^{\rm dq}_{\|\cdot\|,p,d}\cdot\|h\mathbb{1}_{C_k}\|^{\frac1p}_{d/(d+p)}.$$

Sending k to infinity, we get at

$$\liminf_{n\to\infty}n^{\frac1d}\,\bar d_{n,p}(\mathbf{P})\ge Q^{\rm dq}_{\|\cdot\|,p,d}\cdot\|h\|^{\frac1p}_{d/(d+p)}.$$

Step 2 (*Upper bound*, $\mathrm{supp}(\mathbf{P})=\mathbb{R}^d$). Let $\rho\in(0,1)$. Set $K=C_{k+\rho}$ and $K_\rho=C_k$. Let $\Gamma_{k,\alpha,\rho}$ be the lattice grid associated to $K\setminus K_\rho$ with edge $\alpha>0$ as defined in the proof of Proposition 9. It is straightforward that there exists a real constant $C>0$ such that

$$\forall\,k\in\mathbb{N},\forall\,\rho\in(0,1),\forall\,\alpha\in(0,\rho):\quad|\Gamma_{\alpha,\rho}|\le Cd\rho k^{d-1}\alpha^{-d}.$$

Let $\varepsilon\in(0,1)$. For every $n\ge1$, set $\alpha_n=\tilde\alpha_0n^{-\frac1d}$ where $\tilde\alpha_0\in(0,1)$ is a real constant and

$$n_0=|\Gamma_{k,\alpha_n,\rho}|,\quad n_1=\lfloor(1-\varepsilon)(n-n_0)\rfloor,\quad n_2=\lfloor\varepsilon(n-n_0)\rfloor,$$

so that $\alpha_n\in(0.\rho)$, $n_0+n_1+n_2\le n$ and $n_i\ge1$ for large enough n.

For every $\xi\in K_\rho=C_k$, for every grid $\Gamma\subset\mathbb{R}^d$ containing K_ρ, we know by the "firewall" Lemma 2 that

$$F^p(\xi;(\Gamma\cap\mathring{K})\cup\Gamma_{\alpha,\rho})\le(1+\eta)^pF^p(\xi;\Gamma)+(1+\eta)^p(1+1/\eta)^pC_{\|.\|}\alpha^p.$$

Let $\Gamma_1=\Gamma_1(n_1,k)$ be an n_1 quantizer such that $d^p_{n_1}(\mathbf{P}(.|C_k);\Gamma_1)\le(1+\eta)d^p_{n_1}(\mathbf{P}(.|C_k))$. Set $\Gamma_1'=((\Gamma_1\cap\mathring{C}_{k+\rho})\cup\Gamma_{k,\alpha_n,\rho})$. One has $\Gamma_1'\subset C_{k+2\rho}$ for large enough n (so that $\alpha_n<\rho$).

Let moreover $\Gamma_2=\Gamma_2(n_2,k)$ be an n_2 quantizer such that $\bar d^p_{n_2}(\mathbf{P}(.|C_k^c);\Gamma_2)\le(1+\eta)\bar d^p_{n_2}(\mathbf{P}(.|C_k^c))$. For $n\ge n_\rho$, we may assume that $C_{k+2\rho}\subset\mathrm{conv}\,\Gamma_2$ owing to Lemma 1 since $C_{k+2\rho}=\mathrm{conv}(C_{k+2\rho}\setminus C_{k+\frac32\rho})$ and $C_{k+2\rho}\setminus C_{k+\frac32\rho}\subset\overbrace{\mathrm{supp}\mathbf{P}(.|C_k^c)}^{\circ}$. As a consequence $\Gamma_1'\subset\mathrm{conv}(\Gamma_2)$ so that $\mathrm{conv}(\Gamma_1')\subset\mathrm{conv}(\Gamma_2)=\mathrm{conv}(\Gamma)$ where $\Gamma=\Gamma_1'\cup\Gamma_2$and

$$C_{k+\rho}\subset\mathrm{conv}(\Gamma)=\mathrm{conv}(\Gamma_2).$$

Now

$$\begin{aligned}\bar d_n^p(\mathbf{P}) &\le \int_{C_k}\Big(F^p(\xi;\Gamma)\mathbf{1}_{\{\xi\in \mathrm{conv}(\Gamma_2)\}}+\underbrace{d(\xi,\Gamma)^p\mathbf{1}_{\{\xi\notin \mathrm{conv}(\Gamma_2)\}}}_{=0}\Big)d\mathbf{P}(\xi)\\ &\quad+\int_{C_k^c}\big(F^p(\xi;\Gamma)\mathbf{1}_{\{\xi\in \mathrm{conv}(\Gamma_2)\}}+d(\xi,\Gamma)^p\mathbf{1}_{\{\xi\notin \mathrm{conv}(\Gamma_2)\}}\big)\,d\mathbf{P}(\xi).\end{aligned}$$

Using that, for every $\xi\in C_k$,

$$\begin{aligned}F^p(\xi;\Gamma) &\le F^p(\xi;\Gamma_1')\\ &\le (1+\eta)^p\Big(F^p(\xi;\Gamma_1)+(1+1/\eta)^p\,C_{\|.\|}\,\alpha_n^p\Big)\end{aligned}$$

implies

$$\begin{aligned}\bar d_n^p(\mathbf{P}) &\le \mathbf{P}(C_k)(1+\eta)^p\Big((1+\eta)\,d_{n_1}^p(\mathbf{P}(.|C_k))+(1+1/\eta)^p\,C_{\|.\|}\,\tilde\alpha_0\,n^{-\frac1d}\Big)\\ &\quad+\mathbf{P}(C_k^c)\,(1+\eta)\,\bar d_{n_2}^p(\mathbf{P}(.|C_k^c)).\end{aligned}$$

Consequently

$$\begin{aligned}n^{\frac pd}\bar d_n^p(\mathbf{P}) &\le \mathbf{P}(C_k)(1+\eta)^p\Big[(1+\eta)\left(\frac{n}{n_1}\right)^{\frac pd}n_1^{\frac pd}\,d_{n_1}^p(\mathbf{P}(.|C_k))+(1+1/\eta)^pC_{\|.\|}\tilde\alpha_0\Big]\\ &\quad+(1+\eta)\left(\frac{n}{n_2}\right)^{\frac pd}\mathbf{P}(C_k^c)\,n_2^{\frac pd}\,\bar d_{n_2}^p(\mathbf{P}(.|C_k^c))\end{aligned}$$

which in turn implies, using Proposition 10 for the modulus $d_{n,p}$ and the d-dimensional version of the extended Pierce Lemma (Proposition 2) for $\bar d_{n,p}$,

$$\begin{aligned}\limsup_n n^{\frac pd}\bar d_n^p(\mathbf{P}) &\le \mathbf{P}(C_k)(1+\eta)^p\left(\left(\frac{(1+\eta)^{-p/d}}{(1-\varepsilon)(1-Cd\rho k^{d-1}\tilde\alpha_0^{-d})}\right)^{\frac pd}Q_{\|.\|}^{dq}\|h\mathbf{1}_{C_k}\|_{L^{\frac{d}{d+p}}}\right.\\ &\quad\left.+(1+1/\eta)^pC_{\|.\|}\tilde\alpha_0\right)\\ &\quad+\mathbf{P}(C_k^c)\,(1+\eta)\,C_{p,d}\,\|X\mathbf{1}_{\{X\in C_k^c\}}\|^p_{L^{p+\delta}}\left(\frac{1}{\varepsilon(1-Cd\rho k^{d-1}\tilde\alpha_0^{-d})}\right)^{\frac pd}.\end{aligned}$$

One concludes by letting successively ρ, $\tilde\alpha_0$, η go to 0, $k\to\infty$ and finally ε to 0.

Step 3. (Upper bound: general case). Let $\rho\in(0,1)$. Set $\mathbf{P}_\rho=\rho\mathbf{P}+(1-\rho)\mathbf{P}_0$ where $\mathbf{P}_0=\mathcal{N}(0;I_d)$ (d-dimensional normal distribution). It is clear from the very definition of $\bar d_{n,p}$ that $\bar d_{n,p}(\mathbf{P})\le\frac1\rho\bar d_{n,p}(\mathbf{P}_\rho)$ since $\mathbf{P}\le\frac1\rho\mathbf{P}_\rho$. The distribution $\mathbf{P}_\rho$ has $h_\rho=\rho h+(1-\rho)h_0$ as a density (with obvious notations) and one concludes by noting that

$$\lim_{\rho\to0}\|h_\rho\|_{d/(d+p)}=\|h\|_{d/(d+p)}$$

owing to the Lebesgue dominated convergence Theorem. □

Proof of Proposition 1: Using Hölder's inequality one easily checks that for $0\le r\le p$ and $x\in\mathbb{R}^d$ it holds

$$|x|_{\ell^r}\le d^{\frac1r-\frac1p}\,|x|_{\ell^p}.$$

Moreover, for $m\in\mathbb{N}$ set $n=m^d$ and let Γ' be an optimal quantizer for $d_{m,p}(\mathcal{U}([0,1]))$ (or at least $(1+\varepsilon)$-optimal for $\varepsilon>0$). Denoting $\Gamma=\prod_{i=1}^d\Gamma'$, it then follows from Proposition 4(b) that

$$n^{\frac pd}\,d_n^p\big(\mathcal{U}([0,1]^d)\big)\le n^{\frac pd}\,d^p\big(\mathcal{U}([0,1]^d);\Gamma\big)=m^p\sum_{i=1}^d d^p\big(\mathcal{U}([0,1]);\Gamma'\big)=d\,m^p\,d_m^p\big(\mathcal{U}([0,1])\big).$$

Combining both results and reminding that $Q^{\rm dq}_{\|\cdot\|,p,d}$ holds as an infimum, we obtain for $r\in[0,p]$,

$$\big(Q^{\rm dq}_{|\cdot|_{\ell^r},p,d}\big)^p \le d^{\frac{p}{r}-1}\, n^{\frac{p}{d}}\, d^p_{n,|\cdot|_{\ell^p}}\big(\mathcal{U}([0,1]^d)\big) \le d^{\frac{p}{r}}\, m^p\, d^p_m\big(\mathcal{U}([0,1])\big),$$

which finally proves the assertion by sending $m\to+\infty$. □

6 Concluding remarks and prospects

This result does not complete the theoretical investigations about dual quantization (beyond the existence of optimal dual quantizers in the case $p=1$, left open in [12]): the first one is to elucidate the asymptotic behaviour of the constant $Q^{dq}_{\|.\|,p,d}$ coming out in Theorem 2 as d goes to infinity, most likely by showing that $\lim_{d\to+\infty}\frac{Q^{dq}_{\|.\|,p,d}}{Q^{vq}_{\|.\|,p,d}}=1$. From a practical point of view, is it possible to evaluate the mean dual quantization error induced by an optimal Voronoi quantization grid? An answer to that question would be very valuable for applications since many optimal quantization grids have been computed for various distributions (see *e.g.* [8] for Gaussian distributions).

Many natural questions solved in the optimal Voronoi quantization theory remain open. Among others "Is there a counterpart to the empirical measure theorem for (asymptotically) optimal quantizers?" (see Theorem 7.5, p.96 in [5])? "How does dual quantization behave with respect to empirical distribution of i.i.d. n-samples of a given distribution?". Is it possible to develop an infinite dimensional "functional" dual quantization?

A Numerical results for $\bar{d}_{n,2}(X)^2$

In order to support the heuristic argumentation on the intrinsic and rate optimal growth limitation of the truncation error $\mathbf{P}\big(X\notin C_n\big)$ induced by the extended dual quantization error modulus, we consider the two dimensional random variable

$$X=(W_T,\ \sup_{0\le t\le T} W_t),$$

where $(W_t)_{0\le t\le T}$ is a Brownian Motion.

This example is motivated by the pricing of exotic options, where this joint distribution plays an important role.

Using a variant of the CVLQ algorithm (see [12]) adapted for the dual quantization modulus inside C_n and the nearest neighbor mapping outside, we have computed a sequence of optimal grids together with the squared dual quantization error $\bar{d}_{n,2}(X)^2$ and the truncation error $\mathbf{P}\big(X\notin C_n\big)$.

These results are reported in Table A.

Table 1: Numerical results for the dual quantization X

n	$\bar{d}_{n,2}(X)^2$	$\mathbf{P}\big(X\notin C_n\big)$
50	0.04076	0.01784
100	0.01966	0.00795
150	0.01236	0.00412
200	0.00931	0.00141

Furthermore we see in figure 1 a log-log plot for the convergence of the two rates $\bar{d}_{n,2}(X)^2$ and $\mathbf{P}\big(X\notin C_n\big)$.

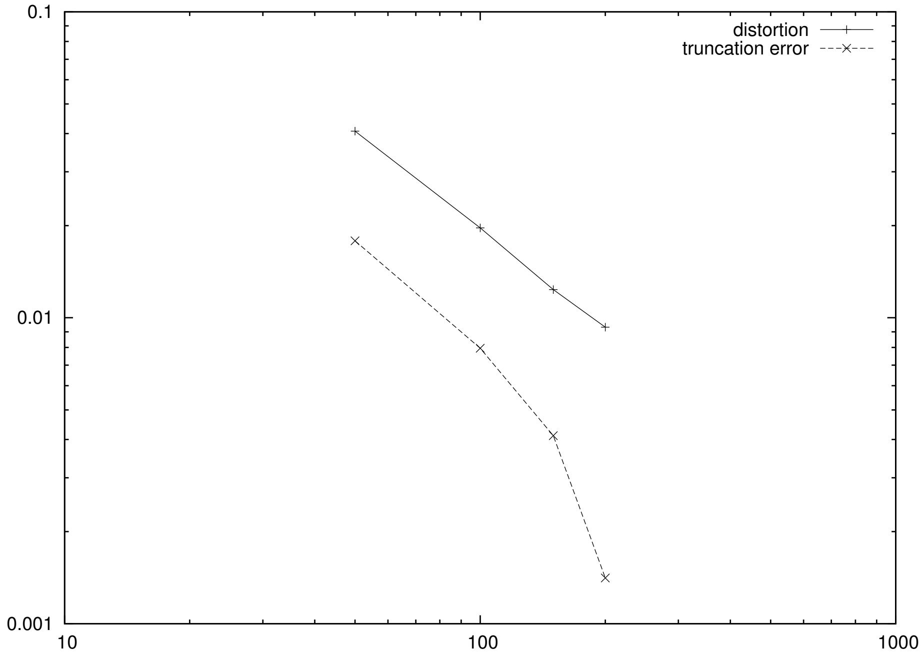

Figure 1: log-log plot of $\bar{d}_{n,2}(X)^2$ and $\mathbf{P}\big(X \notin C_n\big)$ with respect to the grid size n

The distortion rate $\bar{d}_{n,2}(X)^2$ shows here an absolute stable convergence rate (least-squares fit of exponent yields -1.07192) which is consistent with the theoretical optimal rate of $n^{-\frac{2}{d}}$. Moreover, the truncation error $\mathbf{P}\big(X \notin C_n\big)$ outperforms also in this case the heuristically derived rate of n^{-1} and also outperforms the squared "inside" quantization error, which means that also for such an un-symmetric and non-spherical distribution of the Brownian motion and its supremum, an second order rate can be achieved.

This confirms again the motivation of the extended dual quantization error as the correction penalization constraint on growth of the convex hull in order to preserve second order stationarity.

References

[1] J. F. Bonnans, Z. Cen, and T. Christel. Energy contracts management by stochastic programming techniques. *Ann. Oper. Res.*, 200:199–222, 2012.

[2] J. A. Bucklew and G. L. Wise. Multidimensional asymptotic quantization theory with r^{th} power distortion measures. *IEEE Trans. Information theory*, 28(2):239–47, 1982.

[3] P. Capéraà and B. Van Cutsem. *Méthodes et modèles en statistique non paramétrique.* Les Presses de l'Université Laval, Sainte, 1988. Exposé fondamental. [Basic exposition], With a foreword by Capéraà, Van Cutsem and Alain Baille.

[4] P. Cohort. Limit theorems for random normalized distortion. *Ann. Appl. Probab.*, 14(1):118–143, 2004.

[5] S. Graf and H. Luschgy. *Foundations of Quantization for Probability Distributions.* Lecture Notes in Mathematics $n^0$1730. Springer, Berlin, 2000.

[6] R.M. Gray and D.L. Neuhoff. Quantization. *IEEE Trans. Inform.*, 44(6):2325–2383, 1998.

[7] H. Luschgy and G. Pagès. Functional quantization rate and mean regularity of processes with an application to Lévy processes. *Ann. Appl. Probab.*, 18(2):427–469, 2008.

[8] G. Pagès and J. Printems. `www.quantize.maths-fi.com`. website devoted to quantization, 2005. maths-fi.com.

[9] G. Pagès and A. Sagna. Asymptotics of the maximal radius of an L^r-optimal sequence of quantizers. *Bernoulli*, 18(1):360–389, 2012.

[10] G. Pagès and B. Wilbertz. Optimal Delaunay and Voronoi quantization schemes for pricing American style options. in *Numerical Methods for Finance*, R. Carmona, P. Del Moral, P. Hu, N. Oudjane eds, Springer, 2011.

[11] G. Pagès and B. Wilbertz. Dual quantization for random walks with application to credit derivatives. *J. Comp. Finance*, 16(2):33–60, 2012.

[12] G. Pagès and B. Wilbertz. Intrinsic stationarity for vector quantization: foundation of dual quantization. *SIAM J. Numer. Anal.*, 50(2):747–780, 2012.

[13] J. N. Pierce. Asymptotic quantizing error for unbounded random variables. *IEEE Trans. Information theory*, 16(1):81–83, 1970.

[14] V. T. Rajan. Optimality of the delaunay triangulation in $\mathbb{R}^d$. In *SCG '91: Proceedings of the seventh annual symposium on Computational geometry*, pages 357–363, New York, NY, USA, 1991. ACM.